

VINÍCIUS LUIS TREVISAN DE SOUZA

VINICIUS CURALOV

FÁBIO AUGUSTO MORALES MENDES

ALINE HERRERA PEREIRA

CÂMERA 360° 3D

Trabalho de Graduação apresentado à Escola de Engenharia
Mauá do Centro Universitário do Instituto Mauá de
Tecnologia como requisito parcial para a obtenção do título
de Engenheiro de Controle e Automação.

Área de concentração: Engenharia de Controle e Automação.

Orientador: Dr. Eduardo Lobo Lustosa Cabral.

São Caetano do Sul, SP

2017

C189

Câmera 360º 3D / Souza, Vinícius Luis Trevisan de [et al.] - São Caetano Sul :
CEEUN - EEM, 2017.
81 f.

Trabalho de Graduação - Escola de Engenharia Mauá do Centro Universitário
do Instituto Mauá de Tecnologia, São Caetano do Sul, SP, 2017.

Orientadores: Prof. Dr. Eduardo Lobo Lustosa Cabral

1. Visão estéreo 2. Visão 360º 3. Visão computacional
I. Souza, Vinicius Luis Trevisan de II. Mendes, Fabio Augusto Morales .
III. Curalov, Vinícius IV. Pereira, Aline Herrera. V. Instituto de Tecnologia.
Centro Universitário. Escola de Engenharia Mauá. VI. Título.
VI. Título

CDU 621-52

AGRADECIMENTOS

Agradecemos a nossos pais e companheiros por todo o apoio, incentivo e paciência.

Aos nossos professores, sobretudo a nosso orientador, Prof. Eduardo Cabral, por nos guiar no universo da Engenharia.

Aos nossos amigos pela força, ajuda e amizade em todos esses anos.

RESUMO

A indústria de entretenimento tem buscado cada vez mais se reinventar lançando novas formas de consumir mídia. Com a crescente melhoria na capacidade de processamento e memória de dispositivos embarcados, tornou-se viável criar mídias mais complexas, como vídeos em câmera lenta, vídeos de realidade virtual, filmagens 360° ou vídeos em 3D. O propósito deste trabalho é unir duas dessas tecnologias, a capacidade de filmar imagens em 360° e a de criar vídeos em 3D, em um único equipamento, com o intuito de gerar um novo tipo de mídia que cause uma sensação de imersão maior ao usuário. Com isso, o indivíduo assistindo a um vídeo filmado com essa câmera, conseguirá ver o cenário ao redor de si em 360° e manter a sensação de profundidade causada pelas técnicas de imagem 3D. Atualmente existem óculos de realidade virtual que conseguem reproduzir imagens com essas características de imersão, porém, no geral, essas imagens são artificialmente construídas com o uso de um computador. Uma câmera que grava imagens em 360° e 3D permite que óculos desse tipo reproduzam imagens mais reais de shows, festas, eventos e treinamentos.

Palavras-chave: Visão tridimensional, visão 360°, mosaico, visão estereoscópica

ABSTRACT

The entertainment industry has been innovating in delivering new ways to consume media. With the ever growing capacity of the new processors and memories inside embedded systems, it is now possible to create complex types of media, like slow motion videos, virtual reality, 360° videos or 3D videos. The purpose of this work is to unite both the 3D recording and 360° recording technologies into one single device, intended to generate a new kind of media that is able to cause a sensation of immersion to the viewer. Thus, the individual that watches a video made with this camera would be able to see the scenario around him in 360 degrees while still maintaining the sensation of depth caused by 3D imaging techniques. There are, nowadays, virtual reality glasses able to reproduce images and videos with both effects giving the immersion sensation, but generally these images are artificially built with 3D modelling in a computer. A camera that records images in both 360° and 3D allows that real life images, like shows, parties and other events.

Key-words: Tridimensional vision, 360° vision, mosaic, stereo vision.

ÍNDICE DE FIGURAS

Figura 1.1 Linha do tempo de equipamentos de filmagem	13
Figura 1.2 Faturamento do mercado de entretenimento no Brasil em 2015 em milhões de dólares.....	13
Figura 1.3 Projeção de faturamento para o mercado de entretenimento no Brasil em 2020 em milhões de dólares.	14
Figura 1.4 Gasto com entretenimento sobre a renda disponível em 2012.....	14
Figura 3.1 Estereoscópio	17
Figura 3.2 Representação da visão humana.....	18
Figura 3.3 Representação da visão de cada olho	18
Figura 3.4 Representação do campo visual de um humano	18
Figura 3.5 Visão humana e visão estereoscópica com câmeras	19
Figura 3.6 Distâncias interocular e axial	19
Figura 3.7 Sistema TOE-IN	20
Figura 3.8 Sistema de câmeras paralelas	20
Figura 3.9 Panorama da região de North End, Boston	21
Figura 3.10 Panorama criado a partir de 24 imagens	22
Figura 3.11 Relações entre homografias de imagens adjacentes.....	23
Figura 3.12 Projeção planar com câmera em rotação.....	24
Figura 3.13 Projeção cilíndrica.....	25
Figura 3.14 Sistema de coordenadas cilíndricas.....	26
Figura 3.15 Imagem em coordenadas planares	26
Figura 3.16 Imagem em coordenadas cilíndricas	26
Figura 3.17 Mosaico com projeção planar	27
Figura 3.18 Mosaico com projeção cilíndrica	27
Figura 3.19 Mescla por ponto mais próximo do centro.....	28
Figura 3.20 Mescla por média simples.....	28
Figura 3.21 Mescla por feathering.....	29
Figura 3.22 Campo de visão normal.....	29
Figura 3.23 Display esférico.....	30
Figura 3.24 Campo de visão com a lente.....	30
Figura 3.25 Distorção causada pela lente	30
Figura 3.26 Tratamento da distorção	31
Figura 3.27 Projeção da imagem panorâmica Mono	31

Figura 3.28 Projeção imagem Stereo Over Under	32
Figura 3.29 Vuze Camera	33
Figura 3.30 Panocam 3D.....	33
Figura 3.31 360Heros e 360Heros 3D Pro.....	34
Figura 4.1 Vídeos capturados pelos módulos estéreo	35
Figura 4.2- Vídeo cortado	35
Figura 4.3 - Imagens laterais e base dos olhos esquerdos.....	36
Figura 4.4 - Panorâmicas dos olhos direito e esquerdo.....	36
Figura 4.5 - Óculos de realidade virtual.....	36
Figura 5.1 Disposição das câmeras	37
Figura 5.2 Módulo estéreo	37
Figura 5.3 Desenho 3D da câmera	38
Figura 5.4 Desenho da base da estrutura.....	39
Figura 5.5 Desenho da parede lateral.....	39
Figura 5.6 Desenho da parede traseira	39
Figura 5.7 Suporte dos módulos estéreo	40
Figura 5.8 Vista explodida da estrutura	40
Figura 5.9 Foto da câmera montada.....	41
Figura 6.1 Corte de imagem gerada pela câmera.....	42
Figura 6.2 Fluxograma do programa de captura dos vídeos	43
Figura 7.1 Exemplificação do campo de visão da câmera.....	44
Figura 7.2 Processo de captura e reconstrução das imagens.....	44
Figura 7.3 Laboratório de robótica - Imagem base	46
Figura 7.4 Laboratório de robótica - Imagem vizinha esquerda	47
Figura 7.5 Imagem base em coordenadas cilíndricas	48
Figura 7.6 Imagem lateral esquerda em coordenadas cilíndricas	48
Figura 7.7 Cantos de Harris - Imagem base.....	49
Figura 7.8 Cantos de Harris - Imagem lateral esquerda.....	50
Figura 7.9 Cantos de Harris da imagem base após restrições	51
Figura 7.10 Cantos de Harris da imagem lateral esquerda após restrições	51
Figura 7.11 Normalized Cross Correlation entre dois pontos determinados	52
Figura 7.12 Realização do NCC entre duas imagens.....	53
Figura 7.13 Pares de pontos gerados pelo NCC.....	54
Figura 7.14 Avaliação de inliers no RANSAC	55

Figura 7.15 Inliers e outliers separados pelo método RANSAC	56
Figura 7.16 Reprojeção de 1 em 0 utilizando a transformação homográfica \mathbf{H}	57
Figura 7.17 Resultado da aplicação do Guided Matching	57
Figura 7.18 Resultado da aplicação de Guided Matching Refinado	58
Figura 7.19 Reprojeção e mescla do panorama	60
Figura 7.20 Região de transição entre as imagens sobrepostas	61
Figura 7.21 Panorama completo	61
Figura 7.22 Fluxograma do programa principal de obtenção das homografias	62
Figura 7.23 Fluxograma da sub-rotina de obtenção da homografia	63
Figura 8.1 Fluxograma da criação de vídeos	65
Figura 9.1 Projeção do vídeo	66
Figura 9.2 Tela do aplicativo com a lista de vídeos	67
Figura 9.3 Tela do aplicativo projetando o vídeo	67
Figura 9.4 Fluxograma do programa de projeção	68
Figura 9.5 Google Cardboard	69
Figura 9.6 Óculos VR Box	69
Figura 10.1 Resultado da pergunta da venda da câmera	70
Figura 10.2 Resultado da pergunta de venda de serviços	71
Figura 10.3 Resultado da pergunta de sublocação de serviços	71
Figura A.1 Obtenção de pares por NCC	77
Figura A.2 Obtenção de homografia pelo método RANSAC	78
Figura A.3 Obtenção de homografia pelo método Guided Matching	79
Figura A.4 Obtenção de homografia pelo método Guided Matching Refinado	80

SUMÁRIO

RESUMO	5
ABSTRACT.....	6
1. INTRODUÇÃO	12
1.1. Motivações e justificativas	12
2. OBJETIVO.....	16
3. FUNDAMENTAÇÃO TEÓRICA	17
3.1. Visão tridimensional.....	17
3.1.1. Visão humana e o 3D estereoscópico	18
3.1.2. Distância interocular e interaxial	19
3.1.3. Sistema de câmeras 3D estereoscópico.....	20
3.2. Imagens panorâmicas	21
3.2.1. Mosaico.....	21
3.2.2. Homografia	22
3.2.3. Obtenção da homografia	23
3.2.4. Projeção cilíndrica.....	24
3.2.5. Mesclagem das imagens sobrepostas.....	27
3.3. Projeção	29
3.4. Vrview.....	31
3.1. Câmeras similares.....	32
3.1.1. Vuze Camera.....	32
3.1.2. Panocam 3D	33
3.1.3. 360Heros	34
4. DESCRIÇÃO DE FUNCIONAMENTO DA CÂMERA.....	35
5. CONSTRUÇÃO FÍSICA DA CÂMERA	37
5.1.1. Montagem da estrutura.....	40

6.	AQUISIÇÃO DOS VÍDEOS	41
6.1.	Captura do vídeo	41
6.2.	Desenvolvimento do programa de captura do vídeo	42
7.	RECONSTRUÇÃO DAS IMAGENS OBTIDAS	44
7.1.	Obtenção das homografias	45
7.1.1.	Transformação da imagem para coordenadas cilíndricas	47
7.1.2.	Obtenção de pontos notáveis	49
7.1.3.	Associação dos pares de pontos notáveis nas duas imagens	52
7.1.4.	Obtenção da homografia com maior correlação utilizando o método RANSAC	54
7.1.5.	Refinamento da homografia utilizando o método <i>Guided Matching</i>	56
7.1.6.	Refinamento da homografia utilizando o método <i>Guided Matching</i> Refinado	58
7.2.	Construção do mosaico	59
7.2.1.	Aplicação da homografia na imagem transformada	59
7.2.2.	Reprojeção e mesclagem das imagens	60
7.3.	Desenvolvimento do programa de obtenção das homografias	62
8.	CONSTRUÇÃO DOS VÍDEOS	64
9.	REPRODUÇÃO DOS VÍDEOS	66
9.1.1.	Programa de projeção	66
9.1.2.	Fluxograma do programa de projeção do vídeo	68
9.1.3.	Óculos VR	69
10.	RECEPÇÃO DO PÚBLICO	70
11.	CONCLUSÃO	72
11.1.	Sugestões para trabalhos futuros	72
12.	REFERÊNCIAS	73
	ANEXOS	76
A.	FLUXOGRAMA DAS SUBROTINAS DE OBTEÇÃO DA HOMOGRAFIA	77

1. INTRODUÇÃO

O entretenimento é uma das indústrias mais importantes do mundo. Segundo a Rede Paranaense de Comunicação (2015) o entretenimento é o terceiro maior mercado do mundo, concentrando diversos nichos como cinema, televisão, música, internet, dentre outros.

Com o avanço da tecnologia, as empresas de entretenimento precisam se manter atualizadas para impedir a perda de possíveis clientes, porém, algumas são dependentes de determinados equipamentos, como câmeras e equipamentos de cena, que possuem alto custo, dificultando e atrasando o processo de atualização da empresa.

Entre os principais afetados por esse problema estão os produtores de vídeos online, um mercado que nos últimos anos teve um crescimento muito acelerado. De acordo com a ComScore (Soares, 2017), o tempo médio de consumo de vídeos online aumentou em 21% (4,2 minutos por vídeo por usuário em 2015 para 5 minutos) no ano passado e o investimento em publicidade no meio digital aumentou 26% em 2016, chegando a R\$11,8 bilhões e com previsão de alta de mais 26% em 2017, totalizando R\$14,8 bilhões (Marchesan, 2017).

Desde 2015 o Youtube aceita vídeos em 360° e já é possível assistir vídeos digitais criados por softwares de edição que são 360° e 3D. Com o intuito de atender à demanda desse público, foi proposta a criação de uma câmera que seja capaz de realizar filmagens em 360° e 3D, algo que até o momento não existe disponível no mercado (Cossetti, 2015).

Essa câmera permite que criadores de conteúdo que dependem de câmeras para realizar seus projetos tenham acesso a uma nova ferramenta, e possam ampliar as possibilidades de criação de conteúdo, atraindo novos clientes e preenchendo a demanda de vídeos nesse estilo.

1.1.Motivações e justificativas

A tecnologia de filmagem de vídeos vem se desenvolvendo cada vez mais com o passar dos anos (Figura 1.1). No início os vídeos eram em preto e branco, depois vieram os vídeos coloridos que foram evoluindo sua qualidade de imagem e resolução. Após algum tempo surgiram os vídeos HD, full HD e com qualidade de imagem 4K tecnologia a qual está sendo inserida em televisões. As últimas tecnologias existentes no mercado são as de filmagem 3D e filmagens 360°.



Figura 1.1 Linha do tempo de equipamentos de filmagem

A câmera 360° 3D desenvolvida nesse trabalho une as duas últimas tecnologias em um único equipamento e destina-se a todas as pessoas consumidoras de conteúdo de imagens disponibilizando uma nova forma de consumo de conteúdo de entretenimento.

A indústria de entretenimento está entre as indústrias que mais faturam na atualidade. No Brasil, conforme dados do artigo publicado na Folha de São Paulo (Oliveira, F., 2016), representados na Figura 1.2 e na Figura 1.3, o mercado de entretenimento deve atingir um faturamento anual de US\$ 48,7 bilhões (cerca de R\$ 169 bilhões) daqui a quatro anos. O que representa um crescimento de 6,4% ao ano no Brasil, enquanto no mesmo período o setor crescerá globalmente 4,4% ao ano.

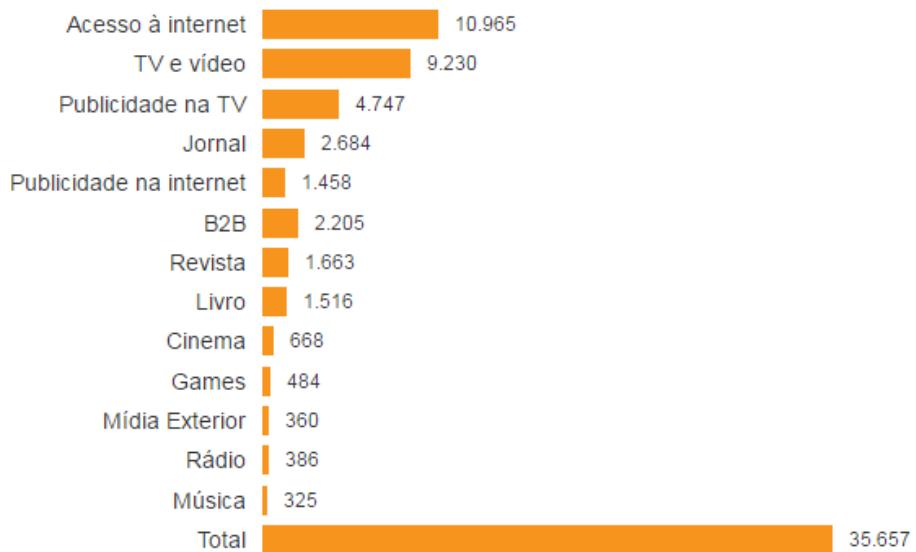


Figura 1.2 Faturamento do mercado de entretenimento no Brasil em 2015 em milhões de dólares.

Fonte: Folha de São Paulo (2016)

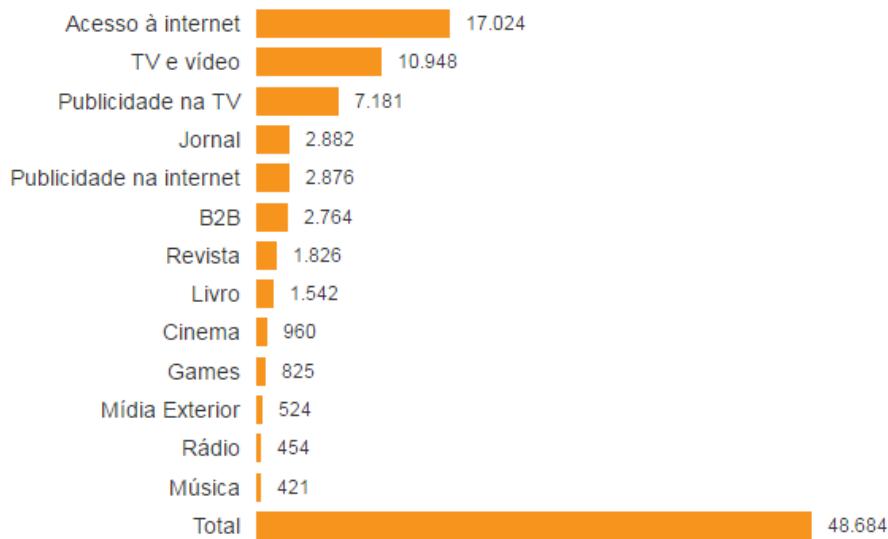


Figura 1.3 Projeção de faturamento para o mercado de entretenimento no Brasil em 2020 em milhões de dólares.

Fonte: Folha de São Paulo (2016)

Segundo dados do estudo publicado no site Time for Fun (2017), o gasto com o entretenimento está diretamente ligado à renda familiar. Os dados estão apresentados na Figura 1.4. Por esse gráfico, nota-se que países como os Estados Unidos, com maior renda per capita, gastam relativamente mais com entretenimento do que países com menor renda per capita, como o Brasil. Porém, pode-se afirmar que mercados de países emergentes, como o do Brasil, Argentina e Chile, tem maior capacidade de expansão do setor de entretenimento.



Figura 1.4 Gasto com entretenimento sobre a renda disponível em 2012

Fonte: Time for Fun (2017)

O site Exame.com (Caleiro, 2015) listou em matéria dois sinais de mudança para a indústria do entretenimento. O primeiro é o fato de o streaming já ter gerado mais receita do que o CD. Números divulgados pela RIAA mostram que em 2014 a venda de música por streaming obteve receita de US\$ 1,87 bilhões, enquanto a venda de CDs físicos teve receita de 1,85 bilhões. Outro sinal é a tendência dos pacotes de TV a cabo começarem a ser desmontados, ou seja, alguns canais terão sua transmissão feita através de app não sendo necessário a compra no pacote de TV. Prova disso são as parcerias como a da HBO e Apple, que disponibilizam conteúdos em streaming e que flexibilizam o acesso a seus conteúdos.

Observa-se que o mercado de entretenimento está em constante mudança, novas tecnologias estão sendo criadas para gerar formas de acesso e produção de conteúdo. Novas tendências fazem parte desse mercado e “ditam” a sobrevivência das empresas no ramo de entretenimento.

A câmera pode ser utilizada para várias aplicações, como por exemplo, festas de casamentos e 15 anos, show, jogos em estádios ou quadras, para fazer um tutorial de um novo veículo que será lançado ou filmar uma viagem de férias. A ideia principal do vídeo é dar uma sensação de imersão no local onde foi filmado.

2. OBJETIVO

O trabalho consiste em projetar e construir uma câmera que grave vídeos unindo os efeitos de filmagem 3D e filmagem em 360° em um único equipamento. As etapas necessárias para o desenvolvimento desse projeto são: construção do módulo de filmagem; desenvolvimento do software de tratamento das imagens e construção do vídeo; percepção do público aos efeitos 360° e 3D, e as possibilidades de mercado.

O princípio básico de funcionamento da câmera 360° 3D é a união de imagens capturadas por diversas câmeras dispostas em torno de um eixo com o intuito de formar um panorama único que permite ao usuário uma visão 360° no entorno do equipamento.

A sensação de profundidade e imersão é criada pela visão estereoscópica. Assim, cada plano de filmagem é capturado simultaneamente por duas câmeras dispostas próximas uma da outra. As câmeras de cada plano que filmam a parte esquerda da cena são as câmeras de olho esquerdo, e analogamente, as que filmam a parte direita da cena são as câmeras de olho direito.

As imagens capturadas pelas câmeras de olho esquerdo de todos os planos são unidas em um panorama, formando uma imagem única, que é em seguida projetada para o olho esquerdo do espectador. Da mesma forma, um panorama é criado para o olho direito, usando as câmeras de olho direito de todos os planos.

A exibição das imagens pode ser feita por um aplicativo de smartphone inserido em um óculos especial, que projeta a imagem de cada olho diretamente para o olho correto do usuário e proporcionando os dois efeitos.

Por limitações financeiras e de tempo, a câmera proposta neste trabalho irá percorrer apenas 180° da cena.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. Visão tridimensional

Existem basicamente cinco métodos de visão tridimensional (3D) nos quais o princípio de formação das imagens são semelhantes (Landim, 2011):

- 3D anáglifo: utiliza um óculos com uma lente azul e a outra vermelha;
- 3D polarizado: duas imagens projetadas são emitidas em polaridades diferentes e um óculos especial separa as imagens direcionadas para cada olho;
- 3D ativo: presente nos óculos que acompanham alguns modelos de TVs 3D;
- Barreira paralaxe: dispensa o uso de óculos e cujo funcionamento é parecido ao dos óculos polarizados, porém todo o processo é realizado por uma barreira colocada na frente da tela;
- 3D estereoscópico.

Neste trabalho o método utilizado de visão 3D é o estereoscópico. Os primeiros registros de fotografias 3D estereoscópicas são datados de 1840. Eram dispostas duas fotos muito idênticas em um aparato, Figura 3.1, a foto da esquerda era vista somente pelo olho esquerdo e a da direita apenas pelo olho direito. O observador olha por meio de duas lentes as fotos quase idênticas e o cérebro combina as duas fotos em uma única imagem em 3D. As técnicas de hoje se assemelham a essas primeiras técnicas desenvolvidas.

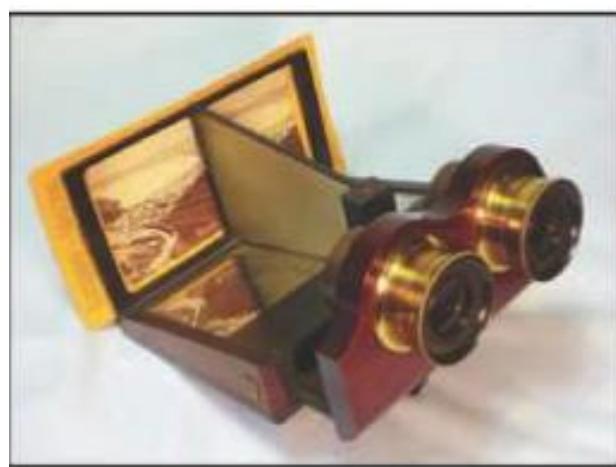
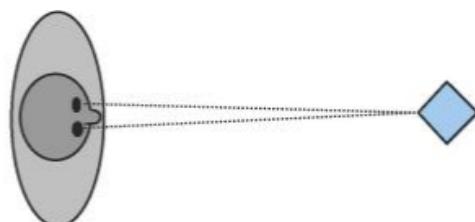


Figura 3.1 Estereoscópio
Fonte: Block (2013)

3.1.1. Visão humana e o 3D estereoscópico

De acordo com Block (2013) a visão normal das pessoas permite que elas vejam o mundo ao redor delas em três dimensões. O mundo é observado por meio de dois olhos, esses convergem sobre os objetos que nos cercam, como ilustrado na Figura 3.2.

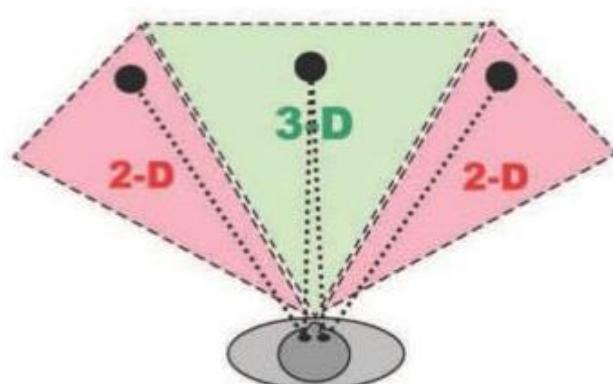


*Figura 3.2 Representação da visão humana
Fonte: Block (2013)*



*Figura 3.3 Representação da visão de cada olho
Esquerda: visão do olho esquerdo. Direita: visão do olho direito. Fonte: Block (2013)*

Os olhos das pessoas estão separados por uma distância de aproximadamente 63,5mm, o que faz com que cada olho enxergue a imagem observada com uma leve diferença, como mostrado na Figura 3.3. Nossa cérebro combina essas duas imagens para formar uma única imagem tridimensional.



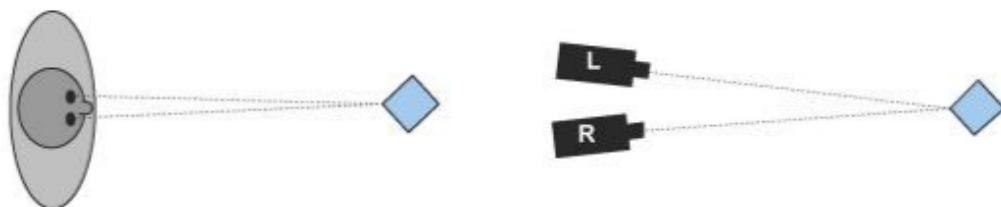
*Figura 3.4 Representação do campo visual de um humano
Fonte: Block (2013)*

A visão humana é 3D estereoscópica somente no centro da área do campo de visão, conforme mostrado na Figura 3.4. A visão do extremo direito do olho direito e vista com o extremo esquerdo do olho esquerdo são vistas de modo 2D. Isso ocorre, pois o nariz interfere no campo de visão de ambos os olhos.

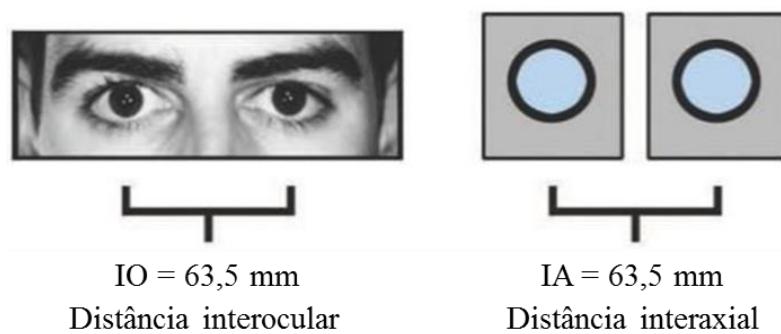
A técnica de visão 3D estereoscópica aplicada em fotografias substitui o par de olhos por um par de lentes de câmeras e isso permite que sejam criadas condições similares ao nosso sistema de visão.

3.1.2. Distância interocular e interaxial

A distância entre os centros dos olhos humanos, de aproximadamente 63,5 mm, é denominada distância interocular (IO). Os olhos humanos em um sistema de filmagem 3D estereoscópico são substituídos por dois pares de lentes (Figura 3.5), e a distância entre os centros dessas lentes é denominado distância interaxial (IA). Essas distâncias estão representadas na Figura 3.6.



*Figura 3.5 Visão humana e visão estereoscópica com câmeras
Fonte: Block (2013)*



*Figura 3.6 Distâncias interocular e axial
Fonte: Block (2013)*

Como os olhos, as duas lentes de câmeras convergem sobre os objetos e têm pontos de vista um pouco diferentes entre si. O par de imagens vistas pelas lentes é chamado de par estereoscópico, conforme representado na Figura 3.3.

3.1.3. Sistema de câmeras 3D estereoscópico

Existem basicamente dois tipos de sistemas de câmeras de 3D estereoscópico, o sistema *toe-in* e o sistema de câmeras paralelas que é o utilizado no projeto.

O sistema *toe-in*, representado na Figura 3.7, é composto de um par de câmeras com um ângulo interno entre elas. As duas imagens formam um par estereoscópico. As linhas laranjas indicam o posicionamento dos sensores de imagem dentro das câmeras.



Figura 3.7 Sistema TOE-IN
Fonte: Block (2013)

A Figura 3.8 representa o esquema do sistema de câmeras paralelas utilizado na Câmera 360° 3D desse projeto. Esse sistema de câmeras é chamado de *Horizontal Image Translation* (HIT) e possui duas câmeras dispostas paralelamente lado a lado. Na ilustração as linhas laranjas representam a posição dos sensores de imagem. O sistema HIT é muito utilizado em animações geradas por computador e cada vez mais produções em *live action*¹ utilizam esse sistema. Esse sistema é o aplicado por conta do tipo de câmera utilizado.



Figura 3.8 Sistema de câmeras paralelas
Fonte: Block (2013)

¹ Em português *ação ao vivo*, é um termo utilizado no cinema, teatro e televisão para definir os trabalhos que são realizados por atores reais, ao contrário das animações.

3.2.Imagens panorâmicas

São consideradas panorâmicas as imagens que representam um amplo campo de visão em uma única imagem. A visão humana, em contraste, é binocular e utiliza de duas imagens capturadas, uma por cada olho, que se sobrepõem e são interpretadas como uma única imagem por nosso cérebro.

Na Figura 3.9 é apresentado um panorama capturado com uma câmera de smartphone que possui o modo “panorâmico”. Nota-se que o campo de visão é mais amplo do que o obtido pelo olho humano.



Figura 3.9 Panorama da região de North End, Boston

Uma imagem em 360° nada mais é que uma representação panorâmica do entorno do elemento que a capturou, e é esse conceito que é abordado neste projeto.

3.2.1. Mosaico

A técnica de mosaico é utilizada para se criar uma imagem panorâmica a partir de diversas imagens individuais. Essa técnica consiste em posicionar as imagens de forma que seus pontos comuns se sobreponham e, com isso, criar uma imagem única com um campo de visão superior ao obtido por uma imagem individualmente.

O procedimento de criação de uma imagem panorâmica por um mosaico de diferentes imagens da mesma cena inicia-se com o registro das diferentes fotos que irão compor o panorama maior. Em seguida deve-se sobrepor as imagens de maneira correta, para que tanto a posição quanto a orientação dos pontos no espaço coincidam em todas as fotos. Por fim, a última etapa é mesclar as imagens sobrepostas, de forma que a transição entre uma foto e a próxima seja o mais natural possível.

Capel (2001) criou o panorama da Figura 3.10 a partir do mosaico de 24 fotos diferentes tiradas da mesma cena por uma câmera rotacionando em torno de si.



*Figura 3.10 Panorama criado a partir de 24 imagens
Fonte: Capel (2001)*

3.2.2. Homografia

Imagens de um mesmo plano vistas a partir de uma câmera executando um movimento arbitrário podem ser capturadas em uma homografia planar. Isso significa que a posição e a orientação de um ponto em uma das imagens podem ser representadas no sistema de coordenadas de outra imagem que o contenha.

Essa relação se dá através de uma matriz de homografia planar (3.1). A matriz de transformação (\mathbf{H}) é obtida a partir de pontos cuja posição é conhecida em ambas as fotos, chamados pontos de calibração. Um ponto no sistema de coordenadas 1, representado por $(x_1, y_1, 1)$ pode ser representado no sistema de coordenadas 0 $(x_0, y_0, 1)$ ao ser multiplicado pela matriz \mathbf{H} , como abaixo:

$$\begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} = \mathbf{H} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (3.1)$$

Escolhendo-se uma imagem, f_0 , como a referência global para as outras, uma foto adjacente f_1 pode ter seus pontos representados no sistema de coordenadas global multiplicando cada ponto pela matriz de homografia ${}^0\mathbf{H}$. Da mesma forma, uma foto f_2 , adjacente a f_1 , pode ter seus pontos representados no sistema de coordenadas de f_1 através da matriz ${}^1\mathbf{H}$ e, consequentemente, no sistema de coordenadas de f_0 aplicando novamente a matriz ${}^0\mathbf{H}$. Assim, uma matriz de homografia ${}^0\mathbf{H}_2$ que faria a transformação diretamente de f_2 para f_0 seria

$${}^0\mathbf{H}_2 = {}^0\mathbf{H} {}^1\mathbf{H} \quad (3.2)$$

A Figura 3.11 ilustra a relação dada pela equação 3.2 para três imagens adjacentes. O processo descrito pode ser aplicado a todas as fotos do mosaico.

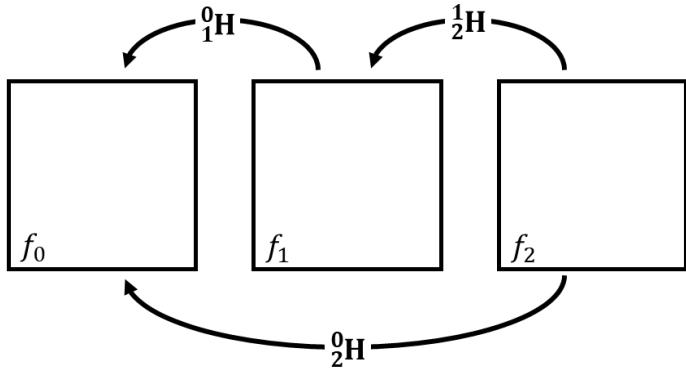


Figura 3.11 Relações entre homografias de imagens adjacentes

Aplicando a homografia pixel a pixel em todas as imagens até que estejam representadas no sistema de coordenadas global, as imagens antes isoladas formam um panorama único.

3.2.3. Obtenção da homografia

A homografia \mathbf{H} é utilizada para correlacionar as coordenadas (x_1, y_1) de um pixel em um sistema de coordenadas de uma imagem com as coordenadas (x_0, y_0) em que esse pixel estaria no sistema de coordenadas de uma outra imagem.

Um ponto (x_1, y_1) no sistema de coordenadas da imagem 1 tem sua projeção no sistema de coordenadas da imagem 0, (x_0, y_0) , definida como:

$$\begin{pmatrix} \lambda x_0 \\ \lambda y_0 \\ \lambda \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (3.3)$$

Pode-se representar a equação 3.3 como um sistema de equações:

$$\begin{aligned} \lambda x_0 &= h_{11}x_1 + h_{12}y_1 + h_{13} \\ \lambda y_0 &= h_{21}x_1 + h_{22}y_1 + h_{23} \\ \lambda &= h_{31}x_1 + h_{32}y_1 + h_{33} \end{aligned} \quad (3.4)$$

Substituindo λ e definindo $h_{33} = 1$, tem-se:

$$\begin{aligned} x_0 &= h_{11}x_1 + h_{12}y_1 + h_{13} - h_{31}x_1x_0 - h_{32}y_1x_0 \\ y_0 &= h_{21}x_1 + h_{22}y_1 + h_{23} - h_{31}x_1y_0 - h_{32}y_1y_0 \end{aligned} \quad (3.5)$$

Cada par de pontos gera duas equações (3.5). A homografia \mathbf{H} possui oito incógnitas que devem ser solucionadas, logo, quatro pares de pontos são suficientes para resolver o sistema de equações 3.6 e obter a matriz de homografia.

Em forma matricial:

$$\begin{bmatrix} x_{1_0} \\ y_{1_0} \\ x_{2_0} \\ y_{2_0} \\ x_{3_0} \\ y_{3_0} \\ x_{4_0} \\ y_{4_0} \end{bmatrix} = \begin{bmatrix} x_{1_1} & y_{1_1} & 1 & 0 & 0 & 0 & -x_{1_1}x_{1_0} & -y_{1_1}x_{1_0} \\ 0 & 0 & 0 & x_{1_1} & y_{1_1} & 1 & -x_{1_1}y_{1_0} & -y_{1_1}y_{1_0} \\ x_{2_1} & y_{2_1} & 1 & 0 & 0 & 0 & -x_{2_1}x_{2_0} & -y_{2_1}x_{2_0} \\ 0 & 0 & 0 & x_{2_1} & y_{2_1} & 1 & -x_{2_1}y_{2_0} & -y_{2_1}y_{2_0} \\ x_{3_1} & y_{3_1} & 1 & 0 & 0 & 0 & -x_{3_1}x_{3_0} & -y_{3_1}x_{3_0} \\ 0 & 0 & 0 & x_{3_1} & y_{3_1} & 1 & -x_{3_1}y_{3_0} & -y_{3_1}y_{3_0} \\ x_{4_1} & y_{4_1} & 1 & 0 & 0 & 0 & -x_{4_1}x_{4_0} & -y_{4_1}x_{4_0} \\ 0 & 0 & 0 & x_{4_1} & y_{4_1} & 1 & -x_{4_1}y_{4_0} & -y_{4_1}y_{4_0} \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \quad (3.6)$$

Resolver o sistema acima significa determinar os parâmetros da matriz de homografia. Observa-se que mais do que quatro pares podem ser utilizados no procedimento acima. Nesse caso, o resultado se ajustaria da melhor forma possível a todos os pontos utilizados ao mesmo tempo.

3.2.4. Projeção cilíndrica

No caso de as imagens serem capturadas a partir de uma rotação da câmera (Szeliski, 2010), pode-se interpretar que os planos de captura são tangentes a um cilindro concêntrico ao centro de rotação da câmera. Essa situação é ilustrada na Figura 3.12 e se assemelha à forma como a câmera deste projeto é construída, com módulos cujos planos de captura estão concêntricos a um centro de rotação.

Ao se projetar as imagens capturadas em uma superfície planar e uni-las, ocorre uma distorção que se acentua quanto mais se afasta do centro do panorama final.

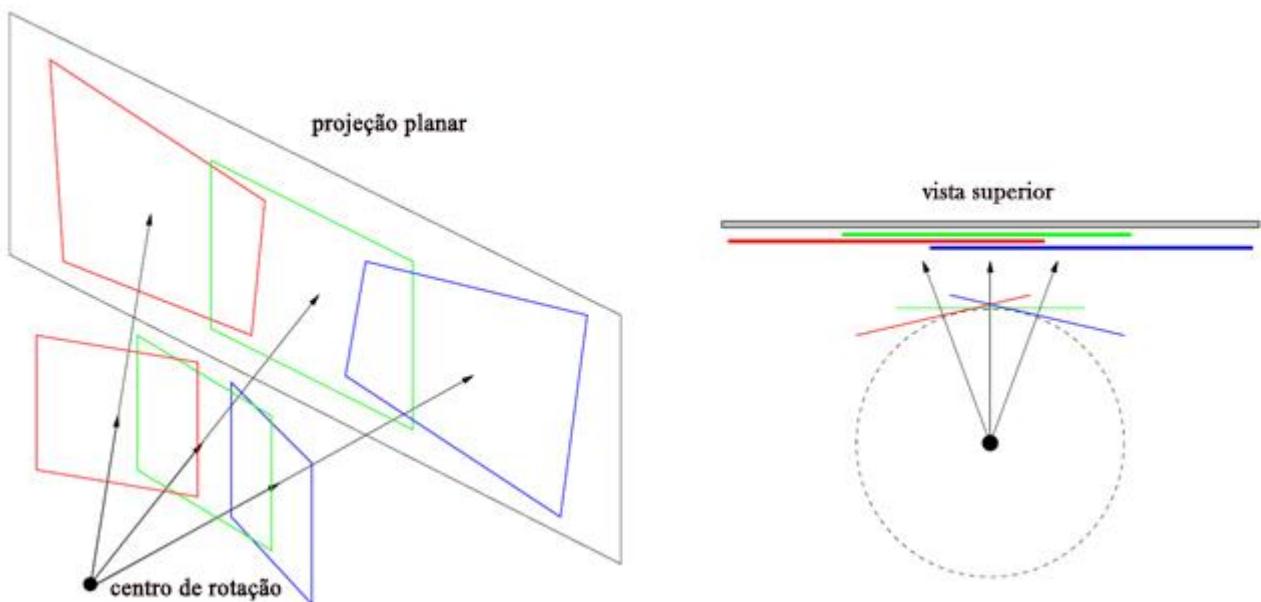


Figura 3.12 Projeção planar com câmera em rotação
Fonte: Capel (2001)

Uma forma simples de corrigir o problema da distorção, conhecido como “efeito gravata borboleta”, por conta do formato distorcido do panorama final, é realizar a projeção das imagens em uma superfície cilíndrica concêntrica ao cilindro de captura, como mostrado na Figura 3.13.

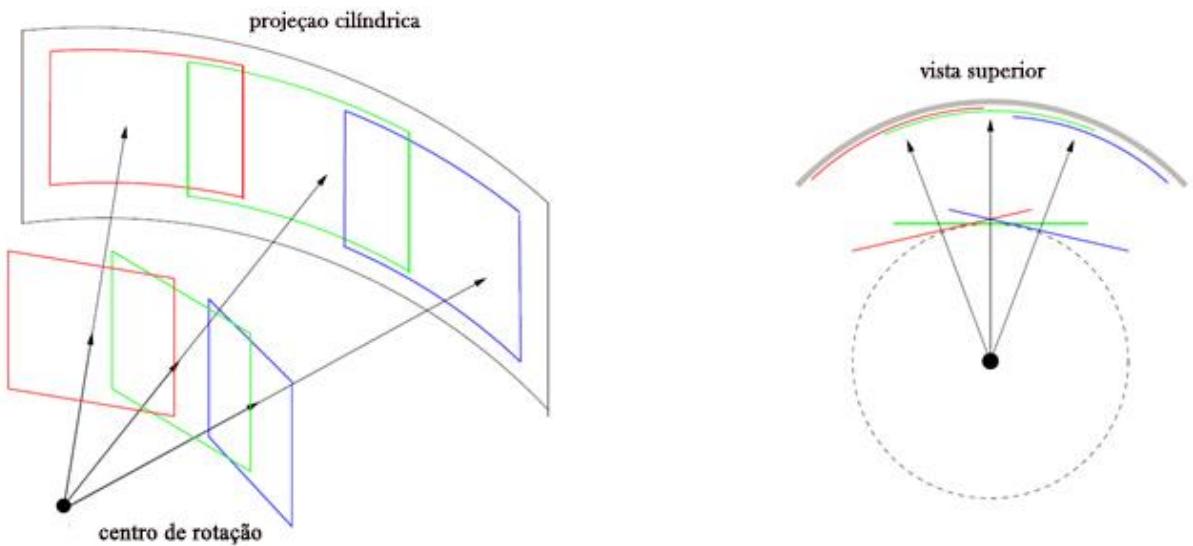


Figura 3.13 Projeção cilíndrica
Fonte: Capel (2001)

Como não há diferença perspectiva entre a superfície de captura e a de projeção, não ocorre o efeito de distorção e a matriz de transformação de coordenadas passa a ser uma simples translação de dois graus de liberdade:

$$\begin{pmatrix} \lambda x_0 \\ \lambda y_0 \\ \lambda \end{pmatrix} = \begin{bmatrix} 1 & 0 & Tx \\ 0 & 1 & Ty \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (3.7)$$

Para se fazer o panorama utilizando projeção cilíndrica deve-se simplesmente transformar as imagens utilizando coordenadas cilíndricas (Figura 3.14). Na figura, um ponto \mathbf{p} de coordenadas (X, Y, Z) é projetado em um cilindro cuja origem (O) do sistema de coordenadas está posicionado em seu baricentro. O ponto \mathbf{x} , que é a projeção de \mathbf{p} , é a intersecção da semirreta \overline{pO} e a superfície do cilindro. A altura entre \mathbf{x} e O é h e a semirreta \overline{Oxp} forma um ângulo θ com a horizontal, portanto, as coordenadas de \mathbf{x} são $(\sin\theta, h, \cos\theta)$.

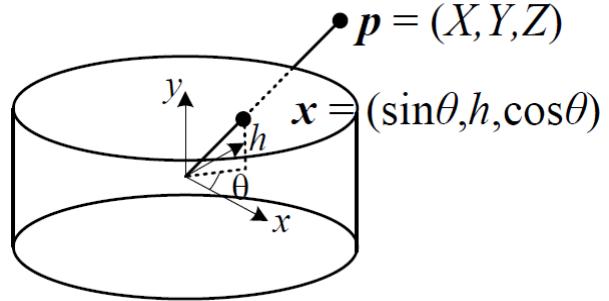


Figura 3.14 Sistema de coordenadas cilíndricas

Considerando uma imagem plana com origem do sistema de coordenadas em seu centro geométrico (x_o, y_o) , qualquer pixel (x, y) dessa imagem será transportado para uma coordenada (x_c, y_c) conforme as equações:

$$x_c = s \tan^{-1} \left(\frac{x}{f} \right) \quad (3.8)$$

$$y_c = s \frac{y}{\sqrt{x^2 + f^2}} \quad (3.9)$$

Onde f é a distância focal do cilindro, utilizada para a projeção, e s é um fator de escala, normalmente definido como sendo igual a f .

Uma imagem em coordenadas planares, como mostrada na Figura 3.15, ao ser projetada cilindricamente, fica como na Figura 3.16, com os extremos superior e inferior arredondados. As retas horizontais da imagem planar se curvam após a transformação, mas as retas verticais se mantém inalteradas.

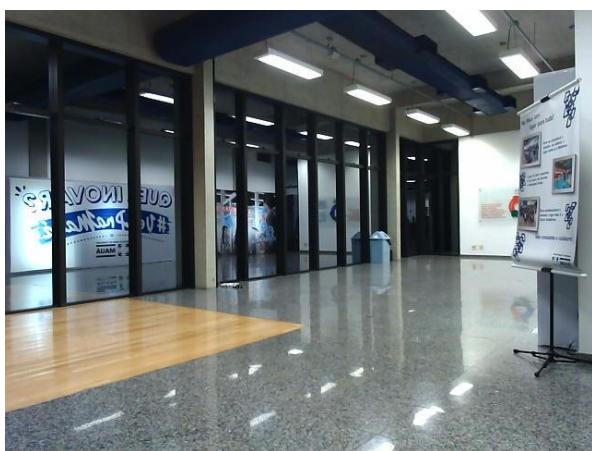


Figura 3.15 Imagem em coordenadas planares



Figura 3.16 Imagem em coordenadas cilíndricas

Um mosaico feito utilizando a projeção planar terá uma distorção do tipo “gravata borboleta”, que aumenta conforme se afasta do centro da imagem em direção aos cantos. Essa situação é ilustrada pela Figura 3.17.

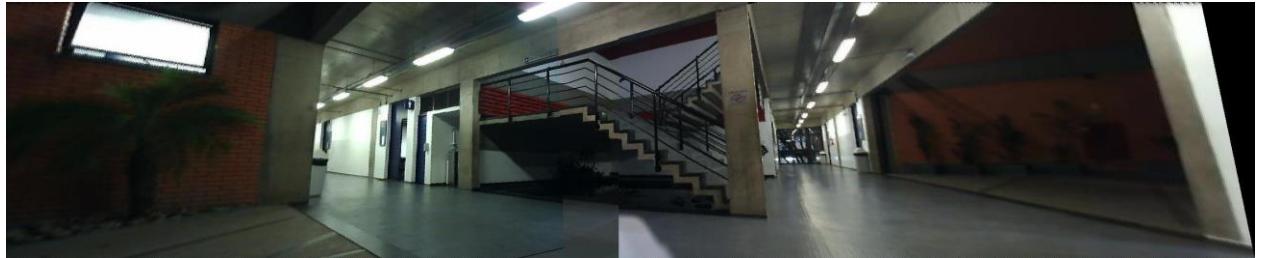


Figura 3.17 Mosaico com projeção planar

Observando cuidadosamente a Figura 3.17 verifica-se que o pilar branco à direita da imagem fica inclinado, e a janela à esquerda é alongada e não apresenta uma altura constante.

Ao se utilizar coordenadas cilíndricas para fazer o mosaico, como na Figura 3.18, a proporção é mantida, porém linhas retas passam a ser curvadas.

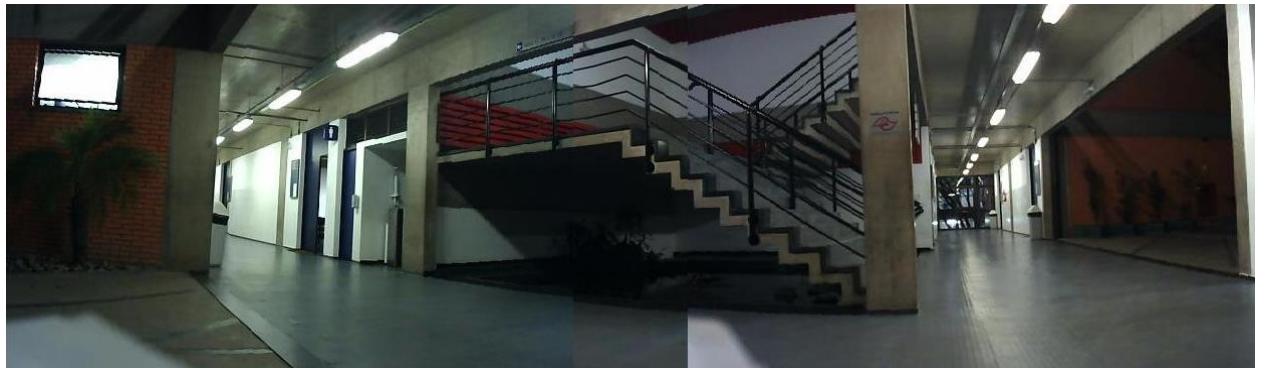


Figura 3.18 Mosaico com projeção cilíndrica

Observando-se a Figura 3.18, em comparação com a Figura 3.17, nota-se que o pilar à direita da imagem se mantém na vertical com suas dimensões inalteradas, e que a janela à esquerda também mantém suas dimensões originais e sua proporção.

3.2.5. Mesclagem das imagens sobrepostas

O resultado do mosaico é um panorama composto de diversas imagens sobrepostas, que nem sempre criam a sensação de continuidade entre as imagens quando dispostas dessa forma. A solução para esse problema é executar uma mescla entre as imagens, de forma que a transição de uma imagem para a outra seja a mais natural possível.

Capel (2001) cita três métodos úteis para a mescla das imagens. O primeiro método é o do ponto mais próximo do centro (Figura 3.19), no qual um ponto que aparece simultaneamente em duas imagens tem sua distância ao centro de cada imagem calculada. O candidato com o menor valor de distância irá compor o mosaico final enquanto o outro é descartado.

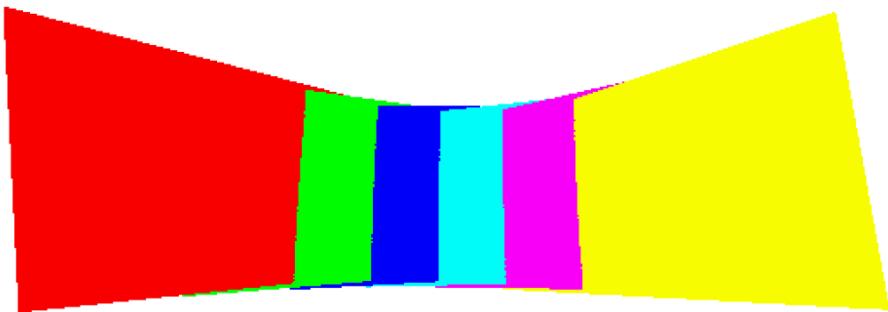


Figura 3.19 Mescla por ponto mais próximo do centro.

Fonte: Capel (2001)

Outro método utilizado é a média simples entre as intensidades das imagens coloridas sobrepostas, visto na Figura 3.20. Onde há sobreposição, os valores de intensidade dos canais vermelho, azul e verde dos pixels são medidos individualmente e então um novo pixel é criado a partir da média dessas medidas.

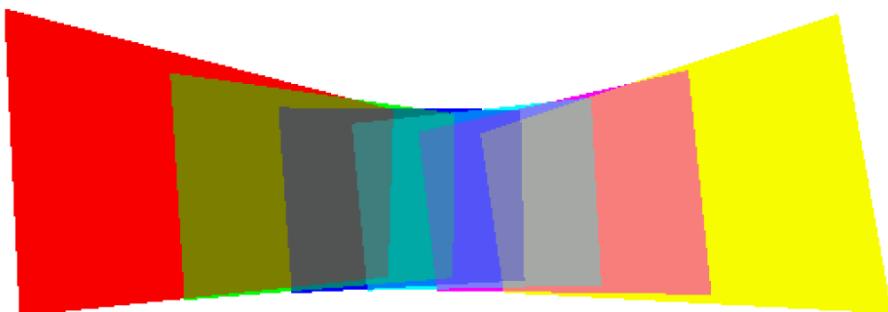


Figura 3.20 Mescla por média simples.

Fonte: Capel (2001)

De forma análoga ao método de média simples, o terceiro método, chamado de *feathering*, é baseado em fazer uma média das intensidades de cada canal de cada pixel individualmente, porém, a média deve ser ponderada por uma função da forma da equação 3.10, que faz com que os pixels mais longe do centro de cada figura tenham um peso (logo, uma influência) menor na imagem final.

$$f(x, y) = \left[1 - \left(x - \frac{w}{2}\right)^2\right] \left[1 - \left(y - \frac{h}{2}\right)^2\right] \quad (3.10)$$

onde x e y são as coordenadas do pixel avaliado no sistema de coordenadas da imagem original e w e h são, respectivamente, a largura e a altura da imagem original, antes de passar pela homografia. O *feathering* está ilustrado na Figura 3.21.

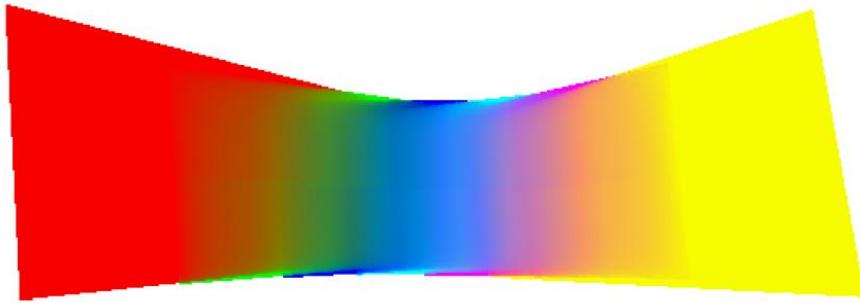


Figura 3.21 Mescla por feathering.
Fonte: Capel (2001)

3.3. Projeção

Quando assistimos a um vídeo ou observamos uma imagem, apenas uma pequena parcela do nosso amplo campo de visão é ocupada, gerando uma zona morta, conforme Figura 3.22, que diminui a imersão do observador.

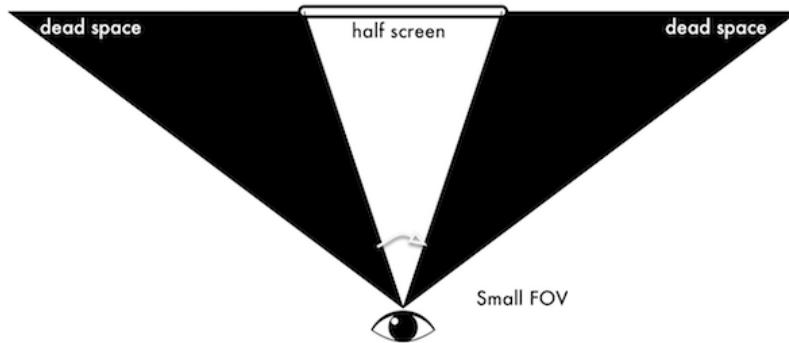


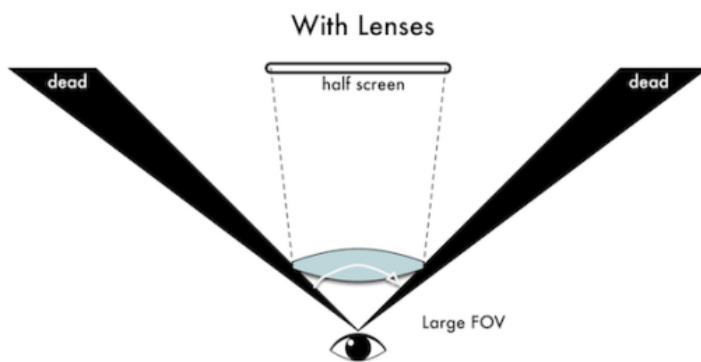
Figura 3.22 Campo de visão normal
Fonte: Boris (2016).

É necessária ocupar uma maior parcela do campo de visão para garantir a imersão no momento da reprodução do vídeo. Esse aumento no campo de visão ocupado pode ser gerado a partir de um grande display esférico, conforme Figura 3.23, próximo ao rosto da pessoa. Por essa tecnologia ser de manufatura complexa, possui um preço muito alto se tornando comercialmente inviável.



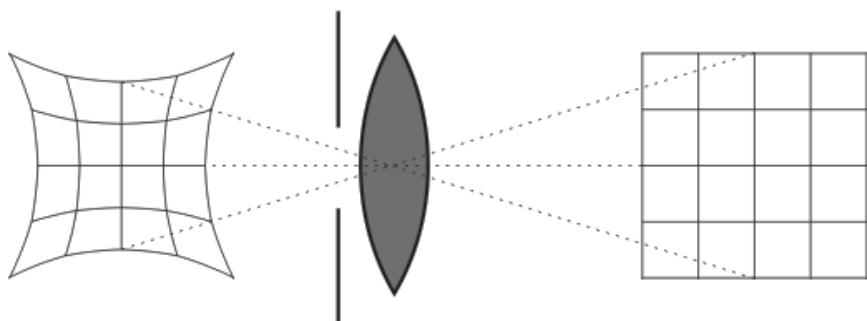
*Figura 3.23 Display esférico
Fonte: Cobra (2014).*

Uma solução mais viável é observar uma imagem plana através de um par de lentes biconvexas, como ilustrado pela Figura 3.24, as lentes causam um efeito de lupa ampliando as imagens e fazendo com que elas ocupem um espaço maior do campo de visão.



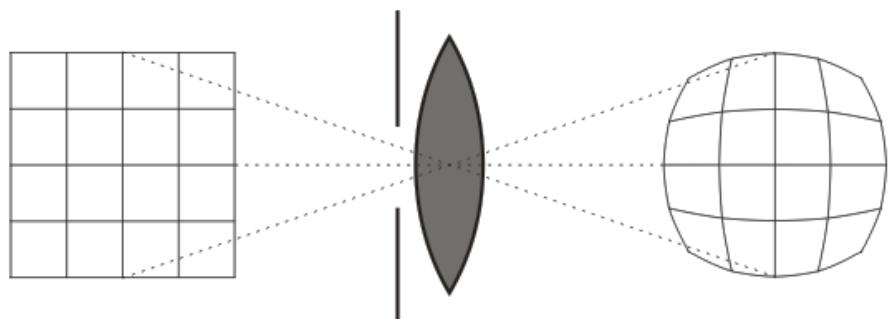
*Figura 3.24 Campo de visão com a lente
Fonte: Boris (2016)*

Conforme mostrado na Figura 3.24, lentes posicionadas próximas ao olho aumentam o campo de visão, mas a imagem sofre também uma distorção esférica proporcional a esse aumento, assim como na Figura 3.25. A forma à esquerda da Figura ilustra a percepção do usuário ao ver através da lente a forma que está à direta.



*Figura 3.25 Distorção causada pela lente
Fonte: Boris (2016).*

É possível corrigir essa deformação na imagem distorcendo-a de modo a compensar o efeito da lente, conforme Figura 3.26. A parte direita da Figura 3.26 é a imagem projetada e a parte esquerda é a imagem vista pelo olho do usuário.



*Figura 3.26 Tratamento da distorção
Fonte: Boris (2016)*

O software de projeção deve fazer esse ajuste na imagem antes de exibi-la ao observador.

Este projeto utiliza o Google VR Java API, que é um conjunto de pacotes fornecidos pelo Google para se programar aplicativos de vídeo 3D.

3.4.Vrview.

Dentro das classes do Google VR Java API, são utilizadas as classes do pacote *com.google.vr.sdk.widgets.video*, especificamente a classe *VrVideoView*.

A classe *VrVideoView* tem como função exibir um vídeo panorâmico, nos formatos Mono ou *Stereo Over Under*, como um vídeo em 3D, quando visto através do Google Cardboard.

Imagens no formato Mono são compostas de uma única imagem panorâmica com uma relação de 2X1 entre o comprimento e a altura, que cria o efeito 3D a partir de dois trechos da imagem com uma distância entre si imitando a visão humana. Cada trecho é projetado em um olho, conforme Figura 3.27.



Figura 3.27 Projeção da imagem panorâmica Mono

Imagens no formato *Stereo Over Under* são compostas por duas panorâmicas retangulares de mesmo tamanho com uma relação de 2X1 entre o comprimento e a altura e alinhadas uma sobre a outra (Figura 3.28). O efeito 3D é criado a partir da projeção de um trecho da imagem de cima projetado no olho direito e um trecho de mesmo tamanho da imagem inferior projetado no olho esquerdo.

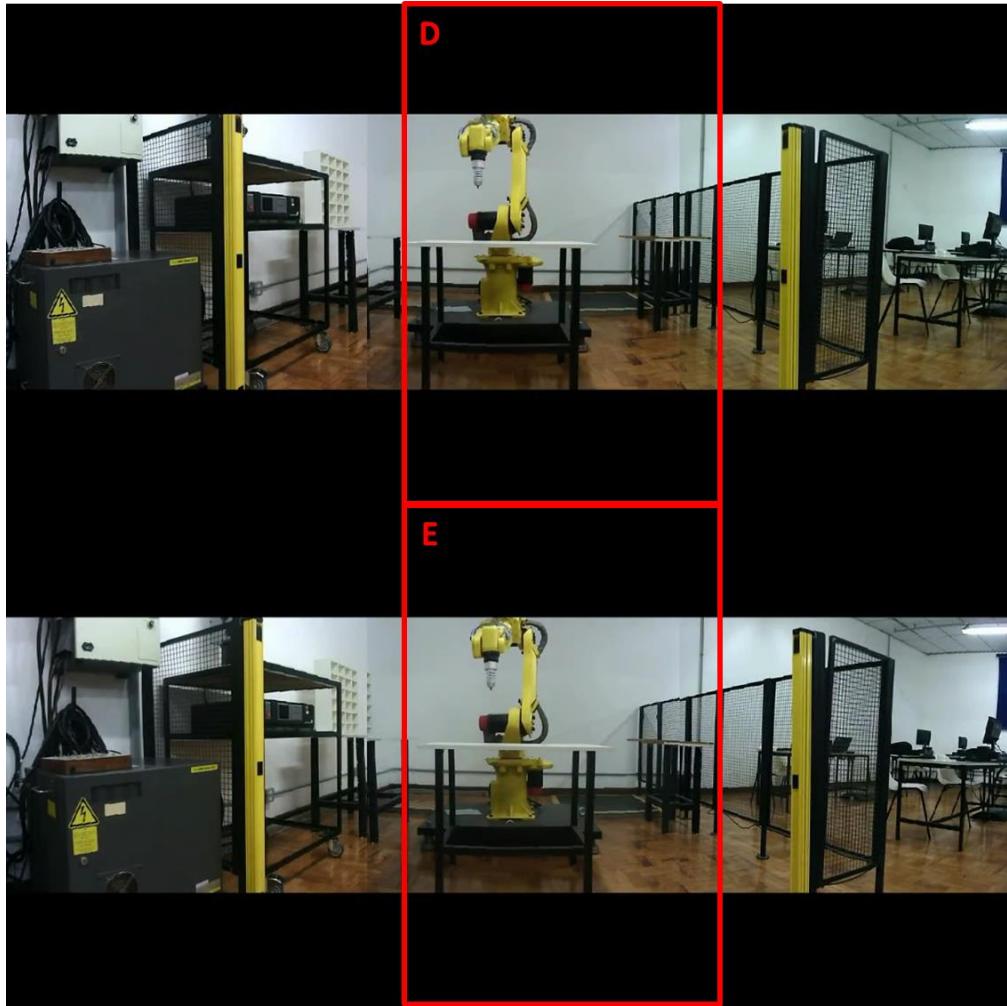


Figura 3.28 Projeção imagem *Stereo Over Under*

3.1. Câmeras similares

Existem, no momento do desenvolvimento deste projeto, câmeras em fase de protótipo que pretendem unir os efeitos de 360° e 3D, porém, nenhuma delas ainda está disponível para venda, sendo apenas protótipos em desenvolvimento. No entanto, muitas delas já estão nos estágios finais de produção e seus fabricantes clamam ter realizado filmagens específicas utilizando-as. Essas câmeras são listadas a seguir.

3.1.1. Vuze Camera

A Vuze Camera, segundo seu site (HumanEyes Technologies, 2017), em tradução livre, “é a primeira câmera de realidade virtual 360° 3D, com alta qualidade e alta resolução, do mundo”. Ela foi criada pela empresa HumanEyes Technologies, que também projetou um aplicativo para smartphones através do qual é possível controlar a câmera e pré-visualizar os vídeos capturados.



Figura 3.29 Vuze Camera
Fonte: HumanEyes Technologies (2017)

3.1.2. Panocam 3D

A Panocam 3D é uma empresa que criou a câmera homônima, conforme visto no site (Panocam 3D, 2017), capaz de criar vídeos em 360° e 3D também a partir da criação de imagens panorâmicas.



Figura 3.30 Panocam 3D
Fonte: Panocam 3D (2017)

No momento a empresa vende apenas os serviços de filmagem, não sendo possível adquirir uma câmera. Para visualizar os vídeos criados com a Panocam é necessário utilizar algum tipo de óculos de realidade virtual.

3.1.3. 360Heros

A 360Heros é um projeto que utiliza 6 câmeras GoPro para criar um sistema que grava as imagens de diferentes planos e as junta, formando uma imagem 360° com todo o entorno da câmera.

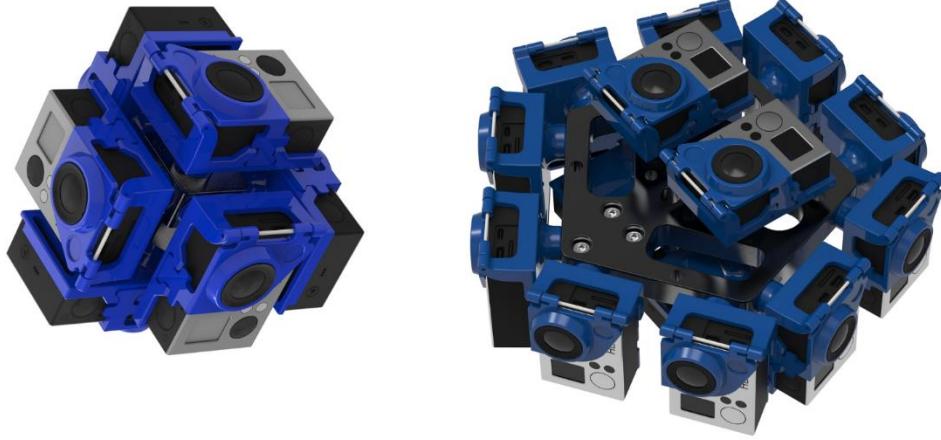


Figura 3.31 360Heros e 360Heros 3D Pro

À esquerda, a 360 Heros, que grava vídeos em 360°. À direita, a 360Heros 3D Pro, evolução da 360Heros que grava vídeos em 360° e 3D. Fonte: 360Rize (2017).

Essa câmera foi utilizada para filmar uma expedição ao topo do Monte Everest, em 2013 e está no Guinness Book com o título de “Primeira filmagem HD de 360° completamente esférica realizada no pico do Monte Everest”.

Uma adaptação no projeto da 360Heros, permitiu a criação de um módulo com o mesmo conceito que cria as imagens em 360° e 3D estereoscópico, conforme visto no site da empresa (360Rize, 2017), a 360Heros 3D Pro, que utiliza 12 câmeras GoPro. Ambos dispositivos estão representados na Figura 3.31.

4. DESCRIÇÃO DE FUNCIONAMENTO DA CÂMERA

O processo geral de funcionamento da câmera pode ser dividido em cinco etapas principais:

- 1) Aquisição de vídeos estéreo: são capturados três vídeos estéreo, um de cada módulo;

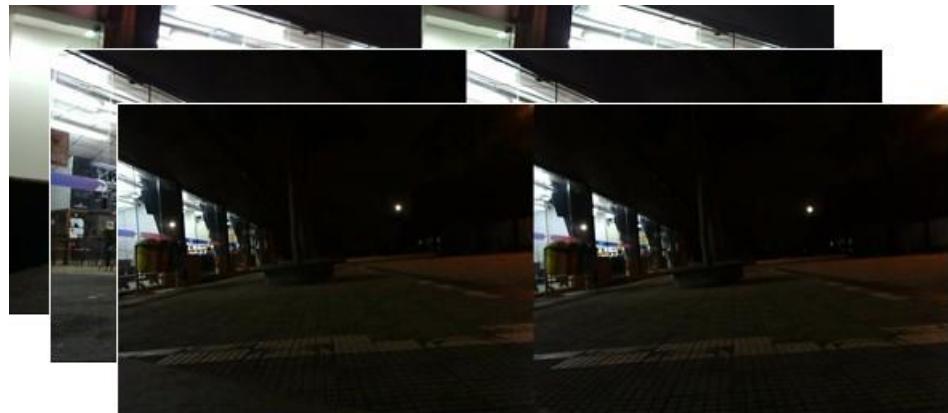


Figura 4.1 Vídeos capturados pelos módulos estéreo

- 2) Corte dos vídeos: os vídeos são cortados separando-se as imagens dos olhos diretos e as imagens dos olhos esquerdos, gerando assim seis vídeos diferentes;

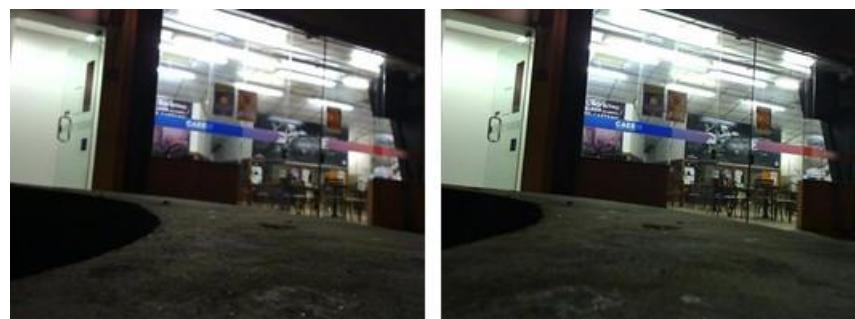


Figura 4.2- Vídeo cortado

- 3) Obtenção das homografias: são obtidas quatro homografias utilizando o primeiro quadro de cada vídeo. As mesmas relacionam, para cada olho, as imagens laterais com a imagem base;

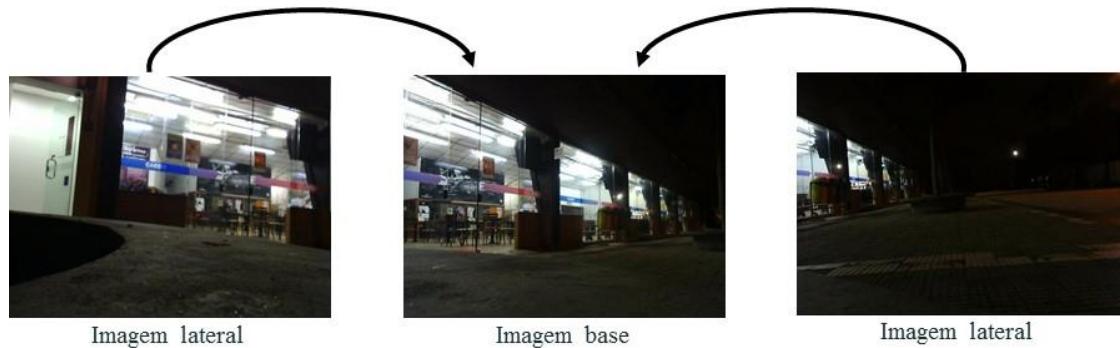


Figura 4.3 - Imagens laterais e base dos olhos esquerdos

- 4) Construção dos vídeos panorâmicos: os vídeos dos olhos direitos são juntados em uma panorâmica e o mesmo procedimento é feito para os olhos esquerdos, obtendo-se duas panorâmicas distintas;

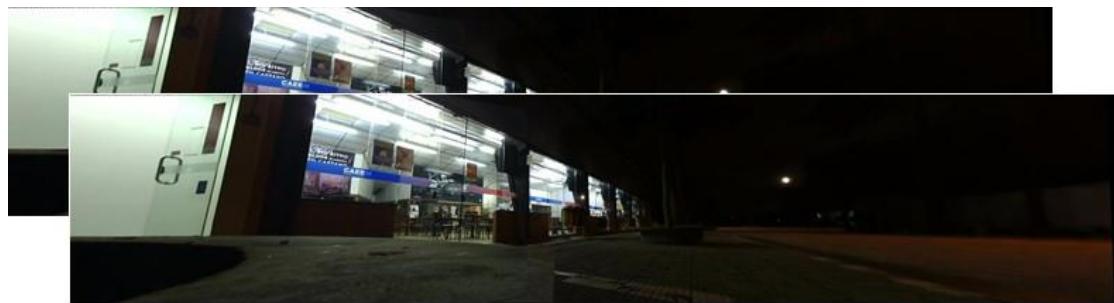


Figura 4.4 - Panorâmicas dos olhos direito e esquerdo

- 5) Exibição nos óculos: as panorâmicas são reprojetadas em um aplicativo para smartphones Android e o vídeo é visto utilizando um óculos de realidade virtual.



Figura 4.5 - Óculos de realidade virtual

5. CONSTRUÇÃO FÍSICA DA CÂMERA

A estrutura da câmera fixa os três módulos de visão estéreo dispostos a 45° , como ilustrado na Figura 5.1. Permitindo que a imagem panorâmica formada pelos módulos estéreo tenha um alcance de 180° .

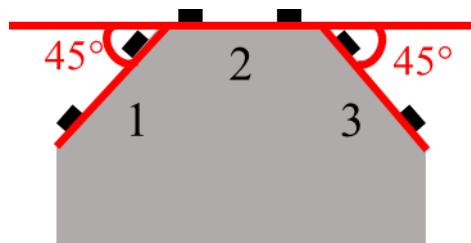


Figura 5.1 Disposição das câmeras

Os módulos estéreo são do fabricante Kayeton Technologies, modelo KYT-U100-960R1ND, como o da Figura 5.2, e possuem uma taxa de captura de 30 fps para uma resolução de 640 x 480 pixels ou de 1280 x 960 pixels. A comunicação é feita com um computador ou hardware similar através de uma porta USB. As lentes são intercambiáveis, para diferentes ângulos de abertura e o foco é regulável manualmente.



Figura 5.2 Módulo estéreo
Fonte: Kayeton Technologies (2017).

A estrutura de fixação das câmeras (Figura 5.3) é feita em acrílico e tem espaço para acomodar três módulos de câmeras, um hardware e os cabos necessário para fazer a ligação com o computador. Ela é composta de quatro tipos de peças: a base, a parede lateral, a parede traseira e o suporte do módulo estéreo.

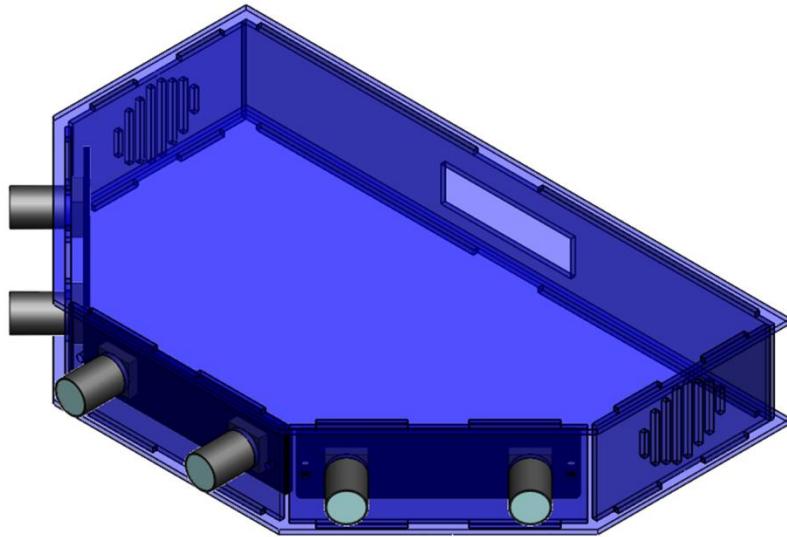


Figura 5.3 Desenho 3D da câmera

As duas bases utilizadas (Figura 5.4) possuem um ângulo de 45° , definido a partir de testes com diversos ângulos, com o intuito de otimizar a quantidade de pontos em comum entre a imagem central e as laterais sem ter muita área de sobreposição entre elas. Os dados do teste estão na Tabela 5.1.

Tabela 5.1 Resultados dos testes de ângulo entre os módulos

α	p	i	Li	Lf	a
30°	29	19	1920	1139	59%
35°	31	21	1920	1244	65%
45°	20	13	1920	1394	73%

onde α é o ângulo entre os módulos, p é a quantidade de pares obtidos pelo método NCC, i é a quantidade de *inliers* obtido pelo método RANSAC, Li é a largura total das imagens dos módulos estéreo, Lf é a largura da panorâmica final e a é o aproveitamento, que é a razão entre Li e Lf .

Observa-se que não há grande variação na quantidade de pares e de *inliers*, mas o aproveitamento aumenta de maneira expressiva.

As bases têm a função de posicionar as outras peças e garantir a estabilidade da montagem.

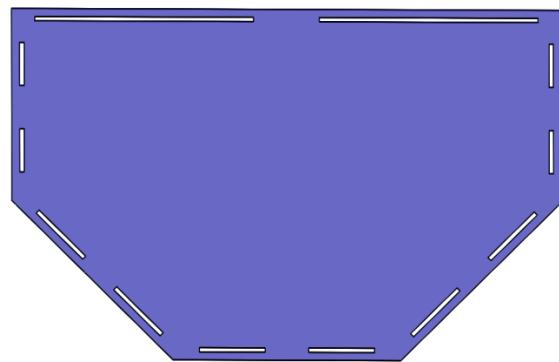


Figura 5.4 Desenho da base da estrutura

A montagem da estrutura da câmera requer a utilização de duas paredes laterais (Figura 5.5) que possuem ranhuras para permitir a circulação de ar por dentro da câmera e evitar o sobreaquecimento dos componentes, quando futuramente houver um hardware embarcado.

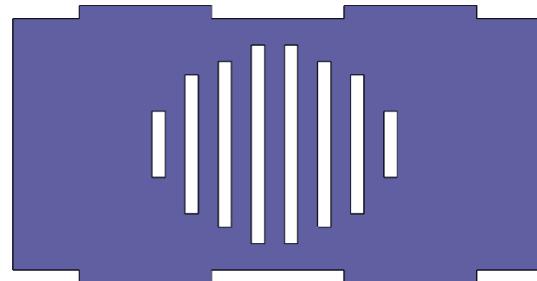


Figura 5.5 Desenho da parede lateral

A parede traseira (Figura 5.6) fornece saída para cabos que fazem a comunicação entre os componentes internos da câmera e aparelhos externos.

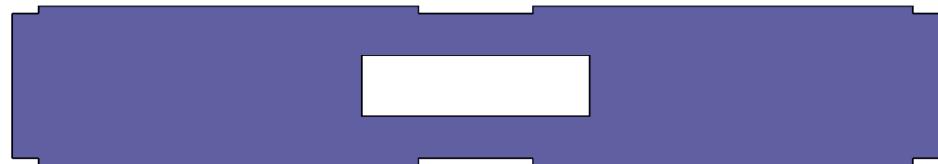


Figura 5.6 Desenho da parede traseira

A estrutura frontal é utilizada para garantir o alinhamento dos módulos, já que para ter uma boa qualidade na união das imagens, todas as câmeras precisam estar alinhadas. O alinhamento e a fixação dos módulos são feitos por parafusos diretamente na estrutura representada na Figura 5.7.

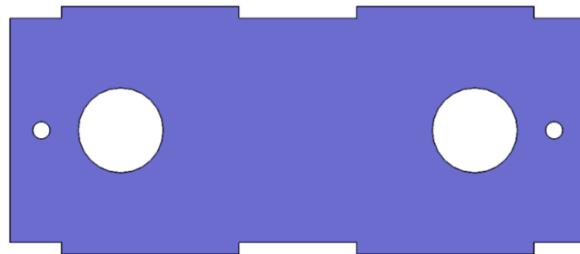


Figura 5.7 Suporte dos módulos estéreo

5.1.1. Montagem da estrutura

A montagem da estrutura da câmera é feita conforme demonstrado na Figura 5.8. Os dentes das paredes laterais e traseira são encaixados nos canais da base da estrutura, as câmeras são parafusadas aos suportes dos módulos e os mesmos presos à base da estrutura por encaixe. A outra base é então utilizada como tampa para garantir a estabilidade da estrutura.

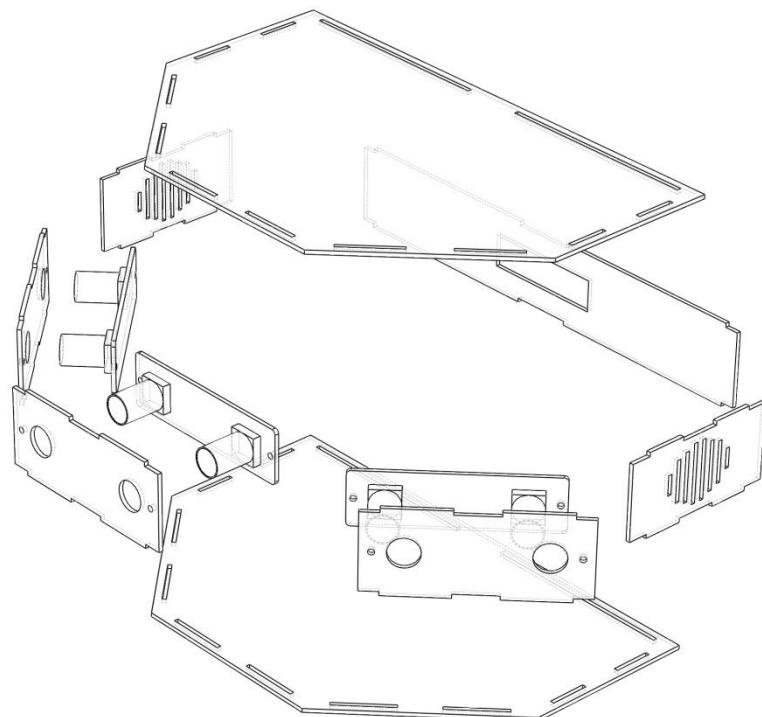


Figura 5.8 Vista explodida da estrutura

A câmera pronta e montada é mostrada na Figura 5.9:



Figura 5.9 Foto da câmera montada

6. AQUISIÇÃO DOS VÍDEOS

A etapa de aquisição consiste na captura do vídeo e na edição (corte e união) dos vídeos de cada módulo de câmeras para formar os vídeos do olho esquerdo e direito.

6.1. Captura do vídeo

A captura das imagens dos módulos de câmeras deve ocorrer de forma simultânea para que seja possível a construção do vídeo a ser mostrado no óculos VR, para tanto os módulos de câmeras devem inicializar e finalizar a gravação do ambiente de maneira conjunta.

Com os três módulos de câmera ligados as portas *USBs* de um computador, pode-se inicializar o programa de captura dos vídeos. O software funciona da seguinte forma, o programa identifica as câmeras conectadas e cria objetos de manipulação dos vídeos, podendo assim, ter o controle simultâneo sobre o tempo de duração dos vídeos. Por meio desses objetos os vídeos também têm suas configurações definidas, como a extensão do arquivo de vídeo, o *codec* de compressão e descompressão, a quantidade de quadros por segundo e a resolução da imagem.

Na etapa seguinte à criação dos objetos de manipulação das imagens, os três vídeos gerados, um de cada módulo de câmeras, são cortados em duas partes, formando assim seis vídeos, conforme Figura 6.1, que mostra o corte das imagens realizado para um módulo de câmeras.

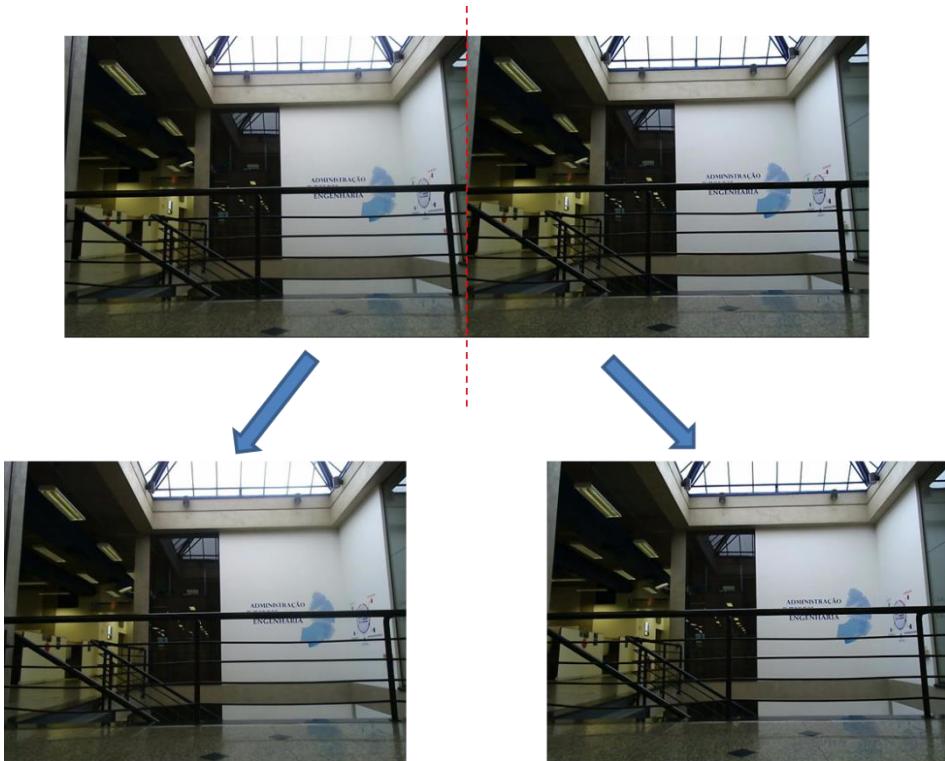


Figura 6.1 Corte de imagem gerada pela câmera

Isso é feito, porque é necessário separar o vídeo da lente esquerda do vídeo da lente direita, para que na etapa seguinte, seja possível juntar as imagens das lentes esquerdas em uma única imagem e as das lentes direitas em outra imagem.

Os seis novos vídeos possuem as mesmas características dos três originais, porém com metade da largura.

Ao terminar a captura, o programa grava os nove vídeos em uma pasta de trabalho, os três vídeos originais, os três vídeos esquerdos e os três diretos de cada módulo de câmera.

6.2. Desenvolvimento do programa de captura do vídeo

O programa foi escrito em Python, utilizando a biblioteca *OpenCV* (OPENCV, 2016).

O fluxograma da Figura 6.2 demonstra o funcionamento do algoritmo de aquisição dos vídeos:

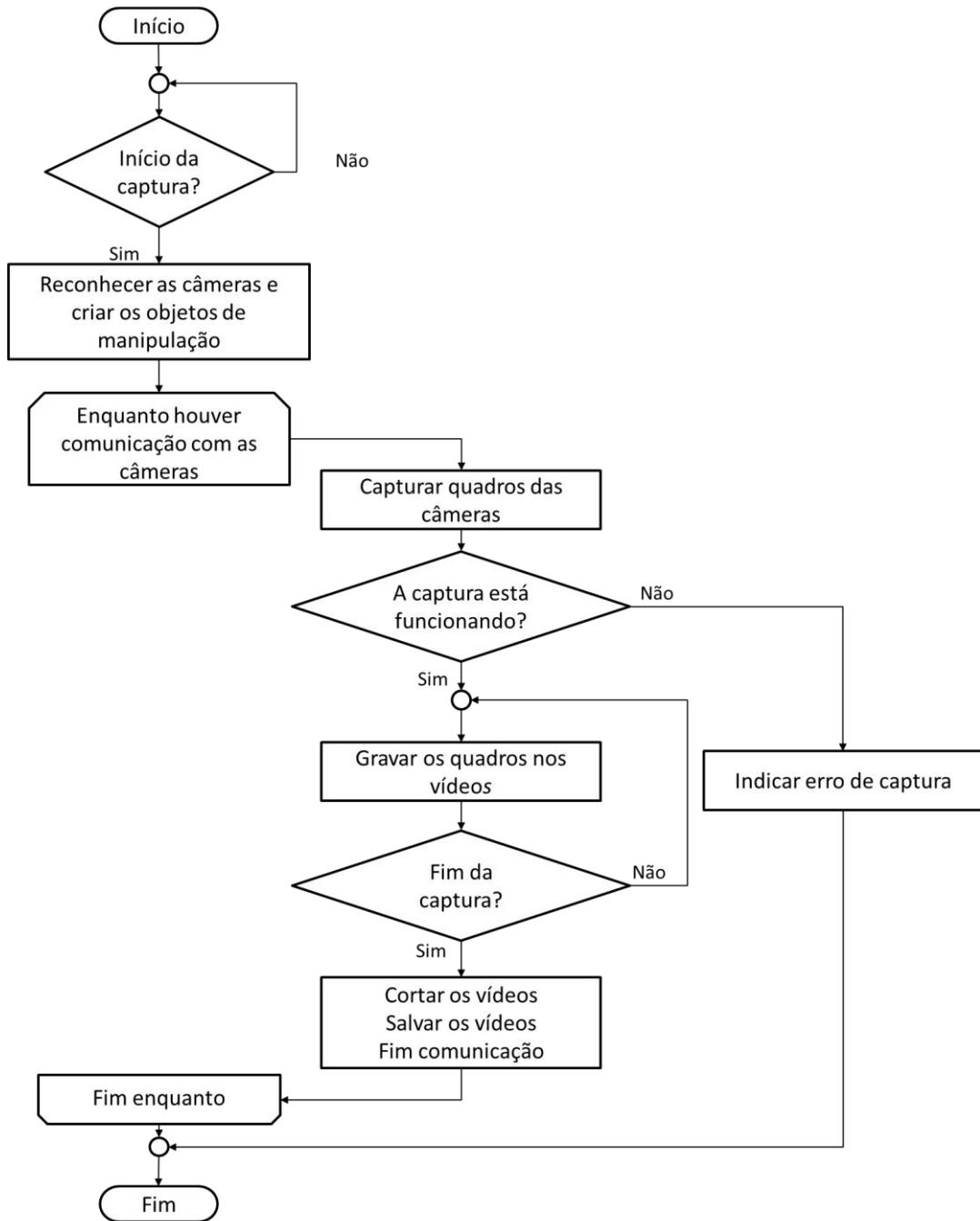


Figura 6.2 Fluxograma do programa de captura dos vídeos

7. RECONSTRUÇÃO DAS IMAGENS OBTIDAS

Cada módulo captura as imagens de seu plano duas vezes: uma com a câmera de olho direito e uma com a câmera de olho esquerdo, conforme mostrado na Figura 7.1.

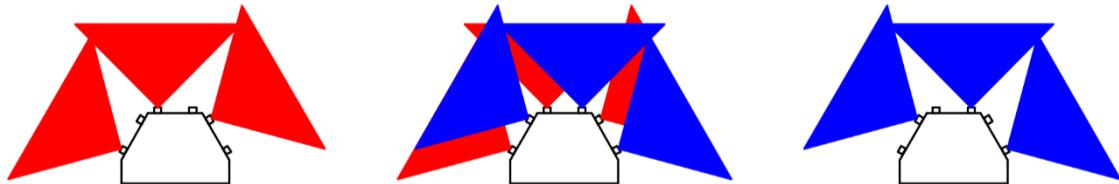


Figura 7.1 Exemplificação do campo de visão da câmera

O esquema do centro da Figura 7.1 mostra a câmera em funcionamento normal e os campos de visão de cada câmera individual do conjunto. À esquerda, em vermelho, apenas o campo visual do olho esquerdo, e à direita, em azul, apenas o campo visual do olho direito.

Todas as imagens capturadas por câmeras de um mesmo olho são unidas em um panorama. Assim, os campos visuais em vermelho irão compor um panorama que é projetado ao olho esquerdo do espectador (Figura 7.2), e da mesma forma, os campos em azul são base para uma imagem panorâmica projetada ao olho direito do usuário.

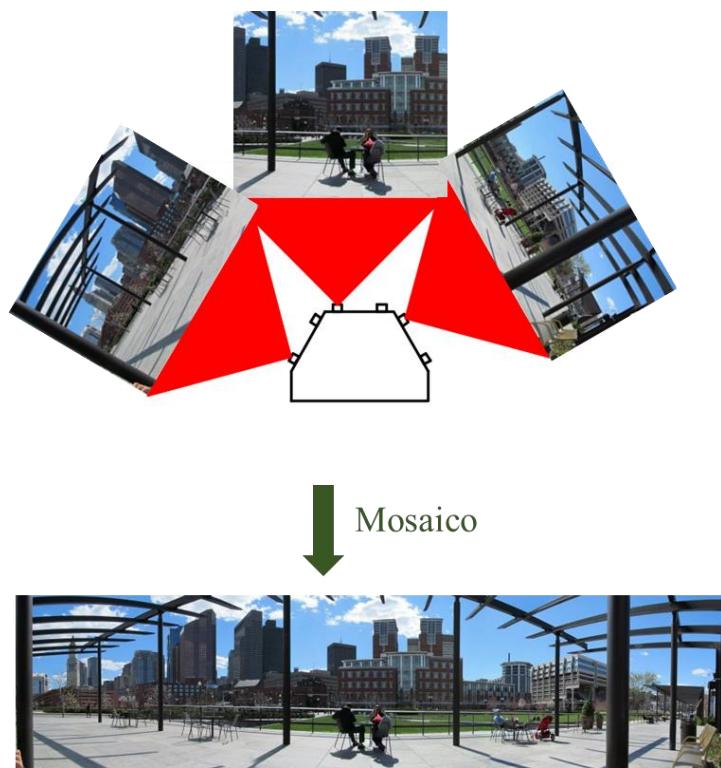


Figura 7.2 Processo de captura e reconstrução das imagens

É importante que as imagens capturadas se sobreponham em algum ponto, para que seja possível encontrar os pontos em comum das duas imagens e, assim, derivar a homografia entre elas.

O objetivo final é construir vídeos utilizando esse processo, assim cada quadro do vídeo consiste na composição das imagens de cada câmera e, ao final, serão montados dois vídeos panorâmicos, um para cada olho.

É necessário obter quatro homografias para realizar o mosaico com os vídeos capturados pela câmera e criar um vídeo panorâmico no formato *Stereo Over Under*. As homografias são:

- H_{le} que faz a relação entre a imagem base do olho esquerdo (img1e) e a imagem esquerda do olho esquerdo (img0e);
- H_{re} que faz a relação entre a imagem base do olho esquerdo (img1e) e a imagem direita do olho esquerdo (img2e);
- H_{ld} que faz a relação entre a imagem base do olho direito (img1d) e a imagem esquerda do olho direito (img0d);
- H_{rd} que faz a relação entre a imagem base do olho direito (img1d) e a imagem direita do olho direito (img2d);

As imagens utilizadas para obter as homografias (img0e, img1e, img2e, img0d, img1d e img2d) são obtidas a partir do primeiro quadro de cada vídeo cortado.

7.1. Obtenção das homografias

De acordo com o mostrado na Figura 7.1 as imagens finais são obtidas do mosaico de seis imagens individuais: três para cada olho. Por exemplo, o olho direito tem uma imagem capturada pela câmera central e uma imagem para cada câmera vizinha.

Para cada imagem lateral ser posicionada no sistema de coordenadas da imagem central é necessária uma homografia diferente. Assim, no total, quatro homografias deverão ser obtidas para a construção dos panoramas, duas para cada olho.

O processo de obtenção das matrizes de homografia aqui ilustrado se aplica para a junção de duas fotos distintas e é utilizado para a obtenção das quatro homografias da câmera.

As etapas são as seguintes:

- Transformação da imagem para coordenadas cilíndricas;
- Obtenção de pontos notáveis;

- Associação dos pares de pontos notáveis nas duas imagens;
- Obtenção da homografia com maior correlação utilizando o método RANSAC;
- Refinamento da homografia utilizando o método *Guided Matching*;
- Refinamento da homografia utilizando o método *Guided Matching* Refinado.

O método aqui utilizado foi baseado no trabalho de Capel (2011) e modificado para as especificidades deste projeto.

A imagem da Figura 7.3 foi capturada com a câmera do projeto em um laboratório do Instituto Mauá de Tecnologia e será utilizada como imagem base para a obtenção da homografia neste tópico. A imagem da figura Figura 7.4 é sua vizinha esquerda. Ambas as imagens possuem 640 pixels de largura por 480 pixels de altura.



Figura 7.3 Laboratório de robótica - Imagem base

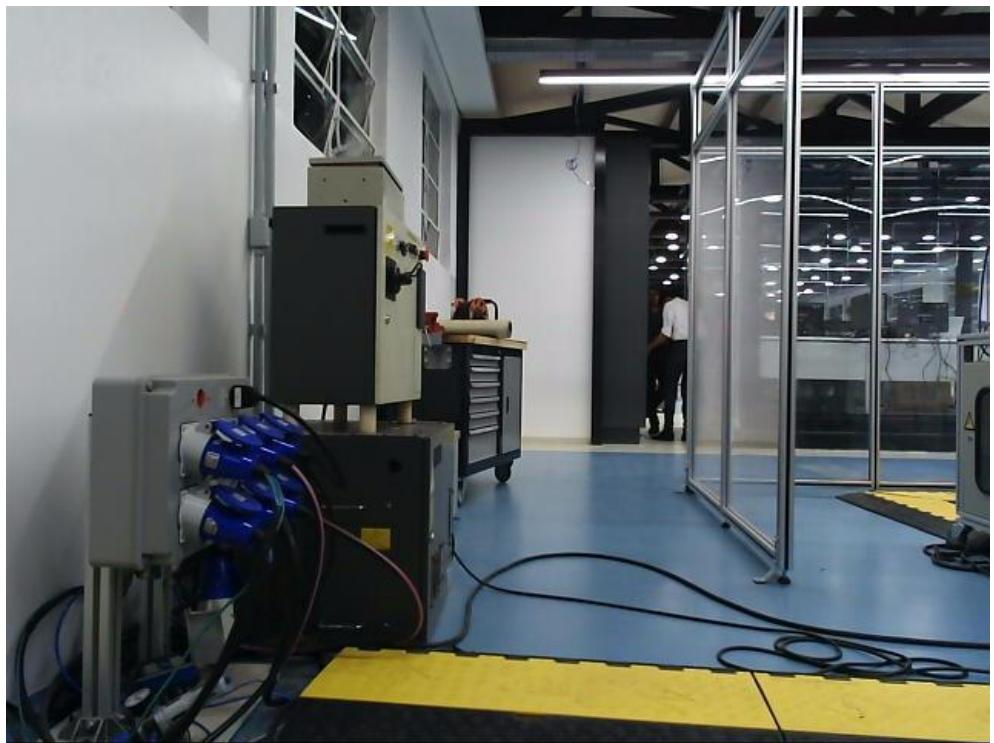


Figura 7.4 Laboratório de robótica - Imagem vizinha esquerda

7.1.1. Transformação da imagem para coordenadas cilíndricas

O método de mosaico por projeção cilíndrica foi escolhido para este trabalho por conta da disposição das câmeras, que estão concêntricas a um mesmo eixo de rotação. Este método permite eliminar a distorção que ocorre nos extremos do panorama que ocorre quando é feita a projeção planar.

Após diversos testes foi escolhido manter a distância focal f como sendo igual à altura da imagem capturada pela câmera, que é de 480 pixels. O parâmetro de escala s também foi definido como sendo igual a f .

Cada pixel de coordenada (x, y) , em relação a uma origem definida como sendo no centro da imagem original, deve ser mapeado como um pixel de coordenada (x_c, y_c) utilizando o método apresentado.

O resultado do mapeamento da Figura 7.3 para coordenadas cilíndricas é apresentado na Figura 7.5 e o resultado do mapeamento da Figura 7.4 é apresentado na Figura 7.6.



Figura 7.5 Imagem base em coordenadas cilíndricas

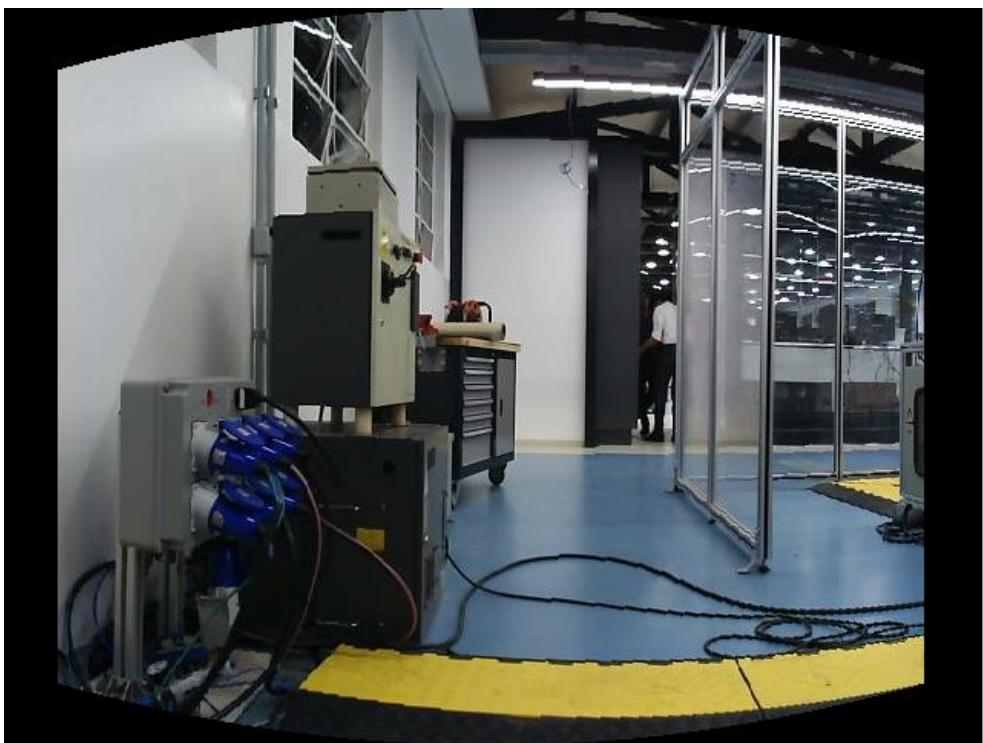


Figura 7.6 Imagem lateral esquerda em coordenadas cilíndricas

7.1.2. Obtenção de pontos notáveis

Uma homografia deve correlacionar pontos no sistema de coordenadas de uma imagem (imagem 1) a pontos no sistema de coordenadas de uma imagem base (imagem 0), portanto, o primeiro passo é a obtenção de pontos notáveis que apareçam em ambas as imagens e sejam de fácil identificação.

O detector de cantos de Harris (Harris, 1988) é o algoritmo utilizado nesse trabalho para essa finalidade. Os cantos detectados na imagem da Figura 7.3 estão representados como círculos verdes na imagem da Figura 7.7 e o resultado para a imagem da Figura 7.4 está representado como círculos vermelhos na imagem da Figura 7.8.



Figura 7.7 Cantos de Harris - Imagem base



Figura 7.8 Cantos de Harris - Imagem lateral esquerda

Em seguida são aplicadas restrições aos cantos obtidos. São desconsiderados os cantos dos extremos superior e inferior das imagens e também os pontos da metade esquerda da imagem lateral esquerda (ou da metade direita caso a imagem seja uma lateral direita), por já estarem fora da área de sobreposição das imagens. O resultado está na Figura 7.9 e na Figura 7.10.



Figura 7.9 Cantos de Harris da imagem base após restrições

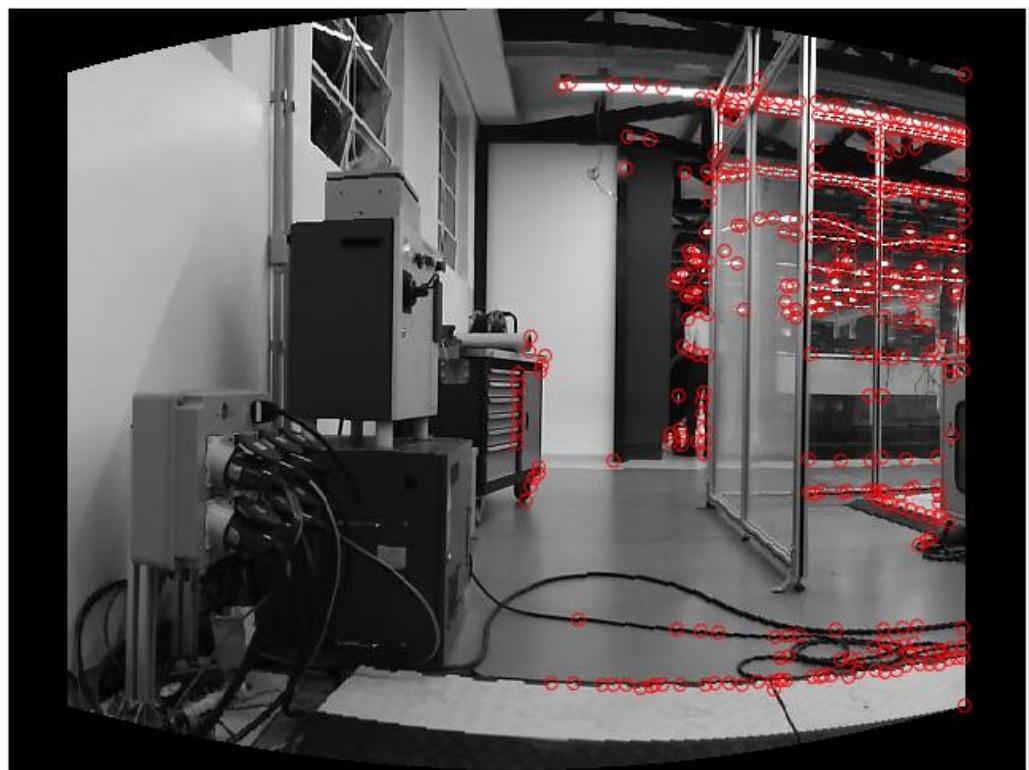


Figura 7.10 Cantos de Harris da imagem lateral esquerda após restrições

7.1.3. Associação dos pares de pontos notáveis nas duas imagens

Cada ponto notável (canto) da imagem 1 deve ser relacionado a um ponto da imagem 0, se possível. Um método computacional então deve ser executado para que os pontos consigam encontrar seus pares na outra figura quando existirem e sejam descartados quando os pares não existirem.

O método utilizado é o de *Normalized Cross-Correlation* (NCC), citado por Capel (2001) e exemplificado por Briechle (2001). Primeiramente seleciona-se um ponto detectado (canto) da imagem 1. Esse ponto é comparado com cada ponto da imagem 0 onde existem cantos e a similaridade entre ambos os pontos é quantificada e calculada. O ponto que tiver a maior similaridade entre todos os comparados será considerado seu par.

Seja $f(x_1, y_1)$ a intensidade do pixel de coordenadas (x_1, y_1) na imagem 1 e $t(x_0, y_0)$ a intensidade do pixel de coordenadas (x_0, y_0) na imagem 0. A correlação γ entre os dois pontos considerando uma janela de largura $2u$ e altura $2v$ ao redor de cada pixel é:

$$\gamma = \frac{\sum_{i,j} (f(x_1+i, y_1+j) - \bar{f})(t(x_0+i, y_0+j) - \bar{t})}{\sqrt{\sum_{i,j} (f(x_1+i, y_1+j) - \bar{f})^2 \sum_{i,j} (t(x_0+i, y_0+j) - \bar{t})^2}} \quad (7.1)$$

onde $i = \{-u, \dots, +u\}$ e $j = \{-v, \dots, +v\}$. O valor \bar{f} corresponde à intensidade média dos pixels da imagem 1 em uma janela de largura $2u$ e altura $2v$ ao redor da coordenada (x_1, y_1) e \bar{t} é o análogo de \bar{f} para a imagem 0. A correlação γ pode variar de 0 a 1.

A Figura 7.11 exemplifica como a comparação é feita. O pixel verde escuro à esquerda está na coordenada (x_1, y_1) da imagem 1 e o pixel vermelho à direita está na coordenada (x_0, y_0) da imagem 0. Os pixels de cor mais clara em volta são a vizinhança considerada no cálculo da correlação γ , ou seja, a janela de $2u \times 2v$ no entorno do pixel central. Para este exemplo foi utilizada uma janela com $u = v = 6$ pixels.

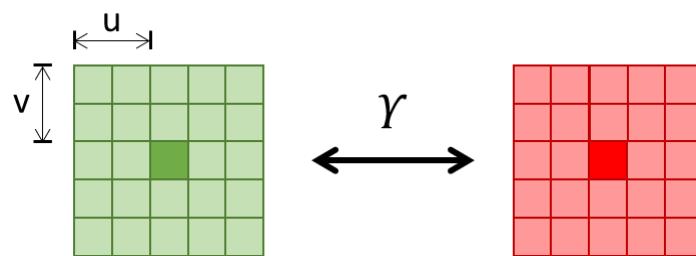


Figura 7.11 Normalized Cross Correlation entre dois pontos determinados

É definido um limiar ϵ de forma que se, para todo ponto de canto da imagem 0 comparado com os pontos de canto da imagem 1, os valores de γ não satisfizerem à equação 7.2 abaixo, o ponto da imagem 1 é considerado um ponto sem par e é descartado. Nesse exemplo foi utilizado $\epsilon = 0,70$.

$$\max(\gamma_k) > \epsilon \quad (7.2)$$

O procedimento é repetido até que o canto escolhido da imagem 1 tenha sido comparado com todos os pontos de cantos da imagem 0 (Figura 7.12). Em seguida o mesmo procedimento é aplicado a todos os pontos restantes da imagem 1 até que todos tenham um par definido ou tenham sido descartados.

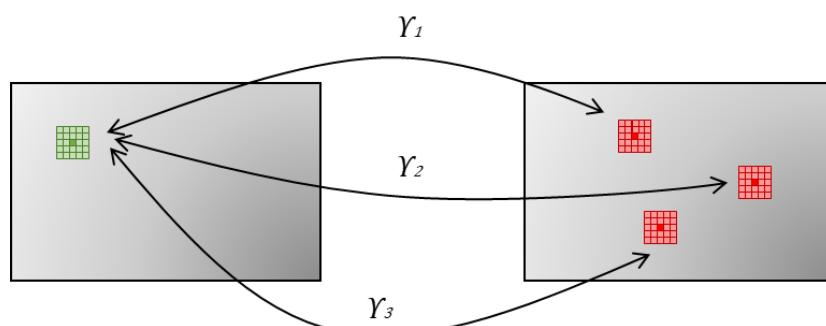


Figura 7.12 Realização do NCC entre duas imagens

Uma restrição de altura foi aplicada para reduzir o espaço de busca e aumentar a eficiência do algoritmo. Caso a distância entre o canto da imagem 1 e o canto da imagem 0 seja maior que um limiar definido, esses pontos não podem formar par.

A Figura 7.13 apresenta a imagem 0 em escala de cinza com seus cantos Harris em círculos verdes e os cantos da imagem 1 sobrepostos como círculos vermelhos. As retas azuis ligando círculos vermelhos aos verdes indicam quais pares foram formados pelo algoritmo do NCC. Os pontos sem retas representam pixels sem pares.

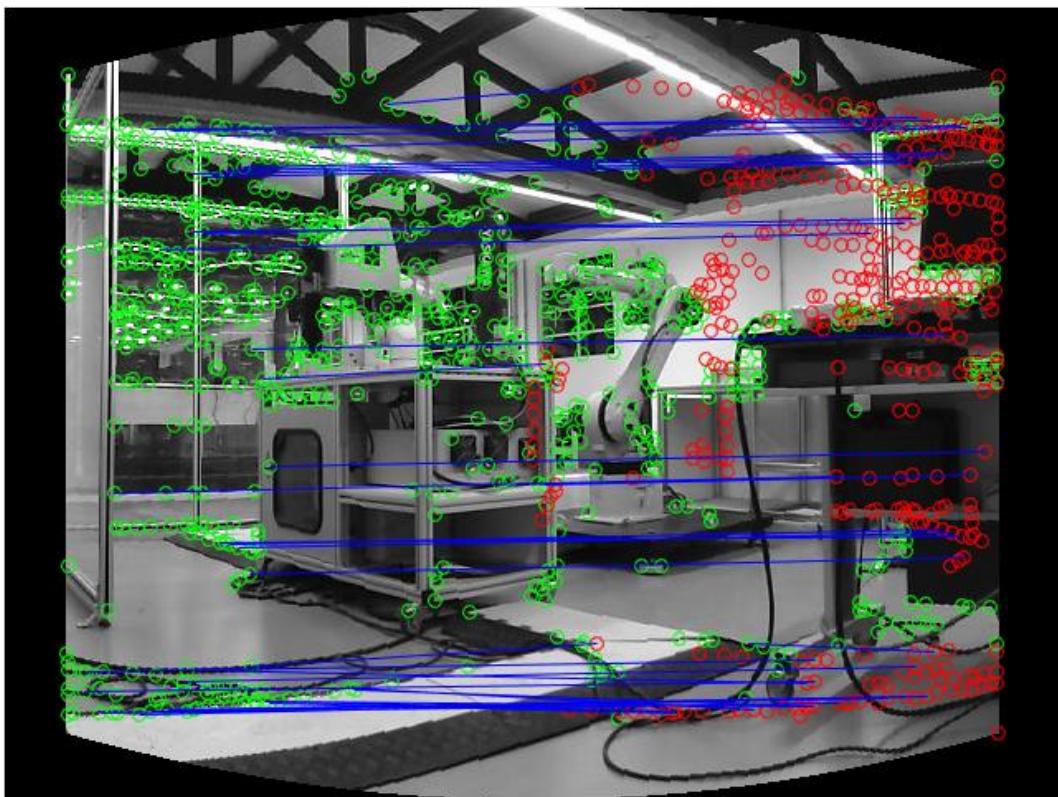


Figura 7.13 Pares de pontos gerados pelo NCC

7.1.4. Obtenção da homografia com maior correlação utilizando o método RANSAC

Observando a Figura 7.13 é difícil inferir quais pares seguem uma mesma orientação e qual das orientações é a mais correta. Nesse sentido, é plausível dizer que esses pares possuem uma grande quantidade de *outliers*, ou seja, pares que não se encaixariam na orientação da maioria.

Um passo importante para a obtenção de uma homografia consistente é a filtragem dessas correspondências para que se possa trabalhar apenas com os *inliers*, os pares que apresentam orientação semelhante à maioria.

O método RANSAC (*RANdom SAmple Consensus*, ou consenso das amostras aleatórias) é aqui aplicado com esse intuito (Fischler, 1981).

A cada iteração são selecionados aleatoriamente quatro pares, que é a quantidade mínima de pares necessários para se realizar o cálculo, e uma homografia é obtida a partir deles. Cada um dos k pontos $f_k(x_1, y_1)$ da imagem 1 que possui um par $t_k(x_0, y_0)$ na imagem 0 é utilizado para calcular a homografia conforme descrito na equação 7.3.

$$\mathbf{r}_k = (x_r, y_r, 1) = \mathbf{H} \mathbf{f}_k \quad (7.3)$$

O ponto \mathbf{r}_k obtido é a projeção do ponto \mathbf{f}_k no sistema de coordenadas da imagem 0 de acordo com a homografia \mathbf{H} . Essa situação é ilustrada pela Figura 7.14.

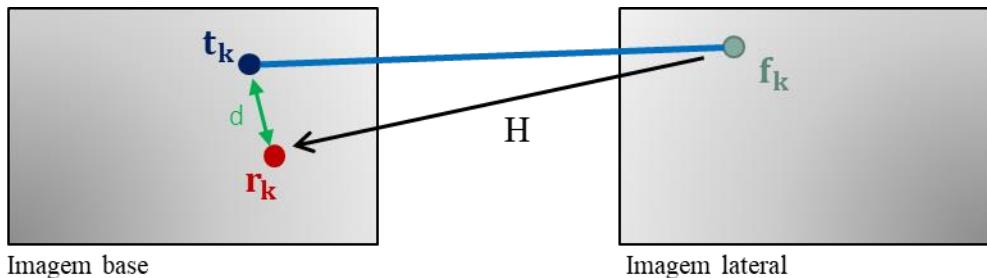


Figura 7.14 Avaliação de inliers no RANSAC

Se o par é um *inlier*, \mathbf{r}_k estará próximo de \mathbf{t}_k considerando-se uma tolerância d :

$$\sqrt{(x_r - x_0)^2 + (y_r - y_0)^2} \leq d \quad (7.4)$$

Se a equação 7.4 não for satisfeita, então o par é considerado um *outlier* para a homografia dessa iteração. Neste exemplo foi utilizado $d = 5$ pixels.

Por conta do caráter aleatório do RANSAC, diversas iterações devem ser executadas com o intuito de obter a homografia que gera o maior número de *inliers*. O algoritmo pode terminar de duas formas: ou com um número fixo muito grande de iterações ou quando a homografia obtiver uma relação *inliers/outliers* favorável (Capel, 2001; Fischler, 1981).

Como para cada imagem a melhor relação *inliers/outliers* possível de se obter é imprevisível, neste trabalho optou-se, após testes, por finalizar o algoritmo após 10.000 iterações.

A Figura 7.15 mostra em azul as retas que representam cada um dos 13 *inliers* e em vermelho as retas que representam os 19 *outliers*. É possível ver que as retas azuis possuem inclinação e comprimento parecidos.



Figura 7.15 Inliers e outliers separados pelo método RANSAC

Após a obtenção da homografia que gera o maior número de *inliers*, todos os *inliers* são utilizados simultaneamente para calcular uma homografia que otimiza a transformação entre todos eles ao mesmo tempo.

7.1.5. Refinamento da homografia utilizando o método *Guided Matching*

Nesse ponto a homografia obtida já reprojeta os pontos da imagem 1 no sistema de coordenadas da imagem 0 com um erro de reprojeção de poucos pixels. Para reduzir ainda mais o erro é utilizada uma técnica de refino chamada *Guided Matching* (Capel, 2001).

Aplicando-se a homografia \mathbf{H} em um ponto notável (canto) P_1 da imagem 1 obtém-se um ponto P_r , que é sua reprojeção na imagem 0. O ponto da imagem 0, P_0 , que corresponde fielmente ao ponto P_1 , seu par, estará em uma área ao entorno de P_r , por conta do erro de reprojeção ainda existente. Essa situação é ilustrada pela Figura 7.16.

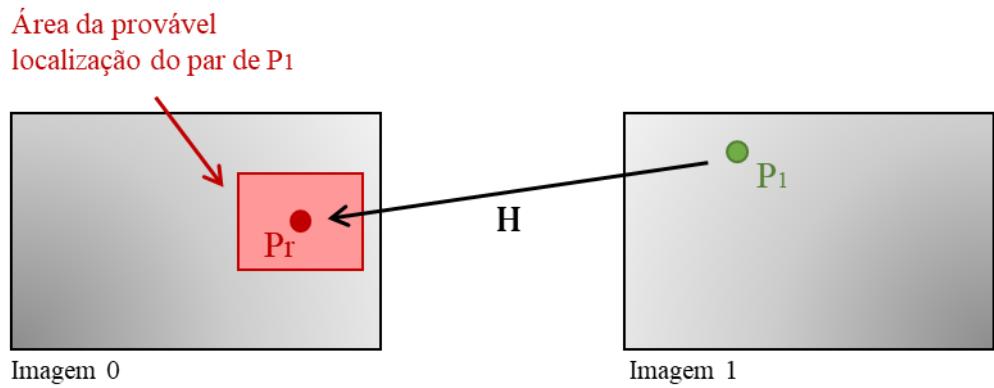


Figura 7.16 Reprojeção de 1 em 0 utilizando a transformação homográfica \mathbf{H}

Para encontrar o par adequado para P_1 basta calcular o NCC entre o ponto P_1 e cada ponto da área que pode conter P_0 (a área ilustrada em vermelho na Figura 7.16). O ponto da área com maior correlação γ que obedece também à equação 6.4 será considerado o par de P_1 .

O resultado da aplicação do *Guided Matching* está ilustrado na Figura 7.17. As retas azuis são os *inliers* obtidos pelo RANSAC e as retas em verde são os novos pares obtidos pelo *Guided Matching*.



Figura 7.17 Resultado da aplicação do Guided Matching

No exemplo, após a aplicação desse método, a quantidade de *inliers* subiu para 20, e uma nova matriz de homografia foi calculada, com menor erro de reprojeção.

7.1.6. Refinamento da homografia utilizando o método *Guided Matching* Refinado

Tomando como base o procedimento da seção 7.1.5 e a matriz de homografia obtida é possível aplicar o *Guided Matching* novamente, mas dessa vez em todos os Cantos Harris da imagem 1, ignorando os pares já formados.

Parte-se do princípio de que uma homografia com pequeno erro de reprojeção pode ser mais efetiva para encontrar um par para um ponto da imagem 1 do que o processo de encontrar os Cantos Harris da imagem 0 e compará-los como feito na seção 7.1.3.

Os pontos da imagem 0 encontrados por esse método, chamado aqui de *Guided Matching Refinado*, se correlacionam com os pontos da imagem 1 de maneira mais fiel do que no método anterior. Além disso, pelo fato da homografia já ter um erro de reprojeção menor, pode-se utilizar uma janela reduzida e um limiar maior para verificar a similaridade usando o NCC.

O resultado da aplicação desse método está na Figura 7.18. Neste exemplo, a quantidade de *inliers* agora é de 51.

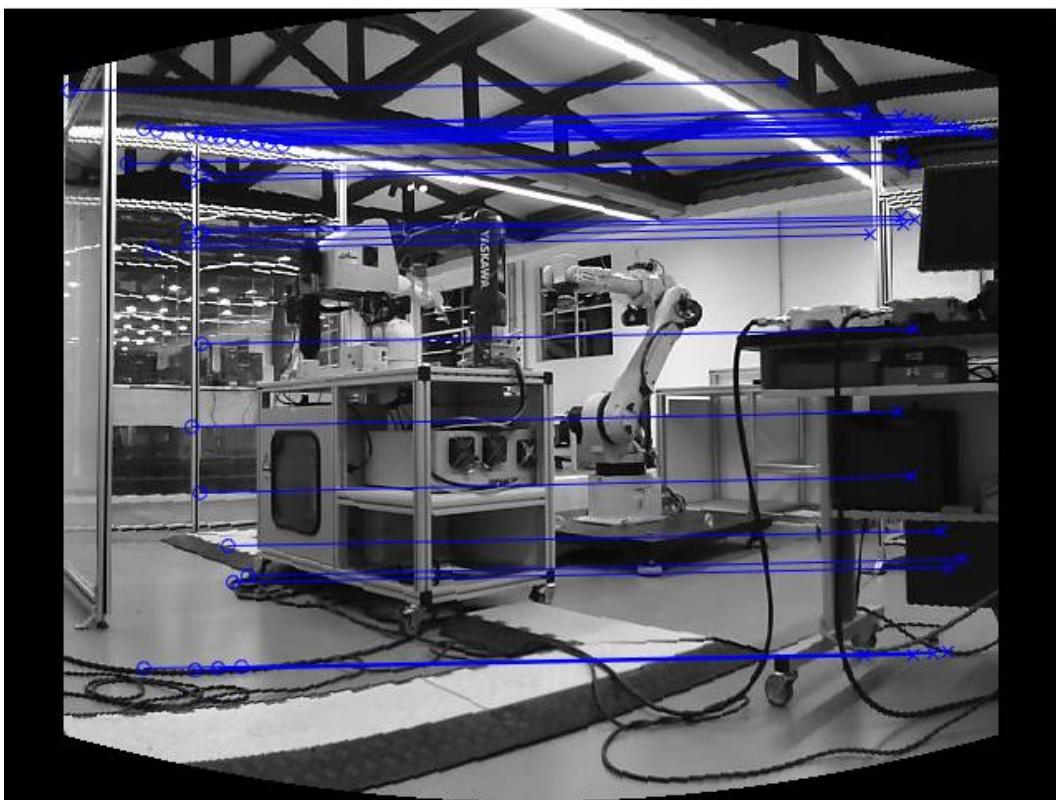


Figura 7.18 Resultado da aplicação de *Guided Matching Refinado*

Uma nova homografia, considerada como sendo o resultado final, é obtida a partir dos novos *inliers*.

7.2. Construção do mosaico

A construção de uma imagem panorâmica a partir de duas imagens diferentes e a matriz de homografia que as relaciona passa pelas seguintes etapas:

- Aplicação da homografia na imagem transformada;
- Reprojeção e mesclagem das imagens.

O procedimento acima deve ser executado para juntar as imagens direita e esquerda com a imagem central, e assim obter uma imagem panorâmica para o olho correspondente da câmera. Em seguida o procedimento é repetido para as imagens do outro olho.

O objetivo final é a obtenção de um vídeo panorâmico. Os vídeos capturados individualmente por cada olho são separados em quadros com cada imagem individual que compõe o vídeo. Em cada quadro as imagens são mescladas e o panorama resultante é o quadro correspondente do vídeo panorâmico. Um vídeo é produzido para cada olho.

7.2.1. Aplicação da homografia na imagem transformada

Para cada pixel da imagem 1 deve-se encontrar seu ponto equivalente no sistema de coordenadas da imagem 0. Seja um pixel de coordenadas (x_1, y_1) da imagem 1, seu correspondente no sistema de coordenadas da imagem 0 é (x_0, y_0) .

A matriz H pode ser representada da seguinte forma:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & Tx \\ 0 & 1 & Ty \\ 0 & 0 & 1 \end{bmatrix} \quad (7.5)$$

Como H é uma matriz 3x3, é necessário acrescentar uma coordenada ao ponto (x_1, y_1) , portanto, ele será reescrito como $(x_1, y_1, 1)$. Como já visto, a projeção desse ponto através de \mathbf{H} é:

$$\begin{pmatrix} \lambda x_0 \\ \lambda y_0 \\ \lambda \end{pmatrix} = \begin{bmatrix} 1 & 0 & Tx \\ 0 & 1 & Ty \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (3.7)$$

Assim, o pixel de coordenadas (x_1, y_1) da imagem 1 original será localizado nas coordenadas (x_0, y_0) da imagem 1 transformada.

7.2.2. Reprojeção e mesclagem das imagens

Diversos métodos de mesclagem de imagens foram apresentados anteriormente e para este projeto o método utilizado foi o de ponto mais próximo do centro.

O motivo disso é que, por conta da transformação cilíndrica, as bordas das imagens apresentam alto grau de distorção. O método de ponto mais próximo do centro faz com que as bordas sejam descartadas, por estarem longe do centro da imagem final. A Figura 7.19 apresenta o panorama após a reprojeção.



Figura 7.19 Reprojeção e mescla do panorama

A transição entre ambas as imagens é suave e a região onde ocorre a transição da imagem base para a transformada é quase imperceptível, como é possível verificar na Figura 7.20 que mostra essa região bem no centro, em destaque.

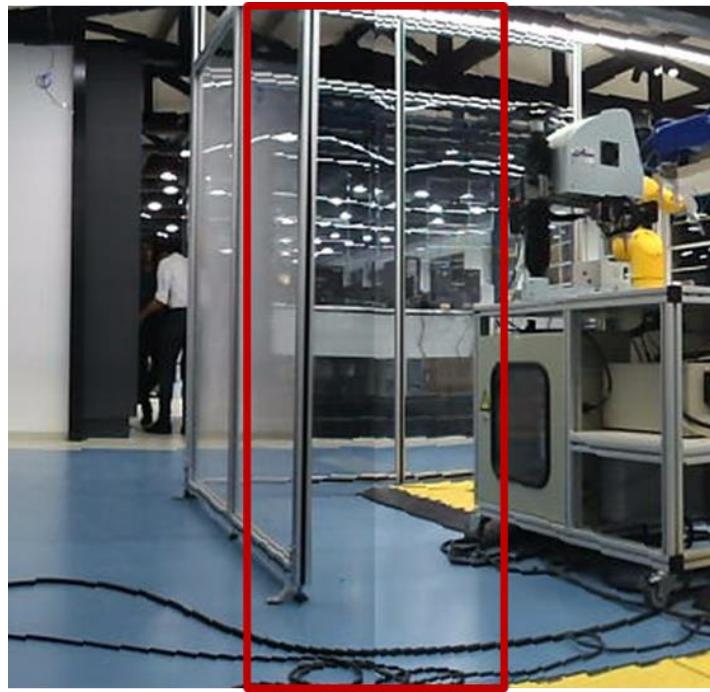


Figura 7.20 Região de transição entre as imagens sobrepostas

Após realizar o mesmo procedimento com uma imagem direita, mesclando todas as imagens em um mesmo panorama, e cortando as bordas, o resultado é a Figura 7.21.



Figura 7.21 Panorama completo

Esse exemplo pode ser considerado um quadro do vídeo de um dos olhos. Os outros *quadros* devem passar pelo mesmo procedimento, formando um vídeo panorâmico completo para cada olho.

Este procedimento deve ser executado mais três vezes até que se obtenha as quatro homografias necessárias.

7.3. Desenvolvimento do programa de obtenção das homografias

O software de obtenção das homografias foi desenvolvido em Matlab, para obter facilmente o retorno visual de cada etapa executada. O programa principal e suas sub-rotinas foram escritos inteiramente para este projeto, mas são utilizadas funções básicas do módulo de processamentos de imagens do próprio Matlab.

Os fluxogramas a seguir ilustram a lógica de funcionamento do software de obtenção das homografias. A entrada do processamento são as imagens comparadas e o resultado é a obtenção das quatro homografias.

A Figura 7.22 ilustra o funcionamento do programa principal.

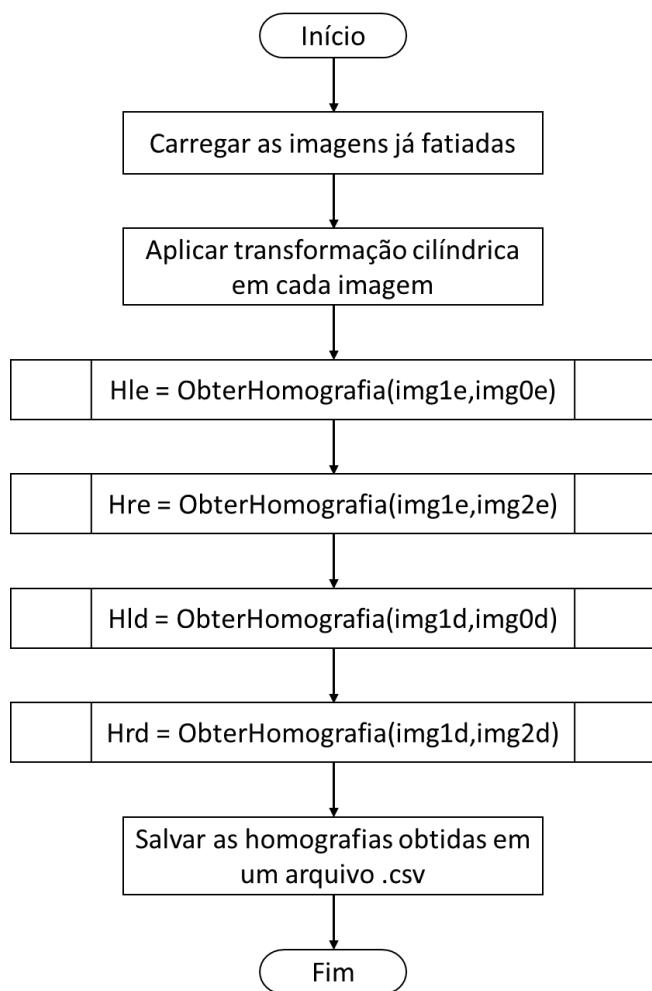


Figura 7.22 Fluxograma do programa principal de obtenção das homografias

A Figura 7.23 apresenta a lógica da sub-rotina de obtenção da homografia. As demais sub-rotinas serão apresentadas em anexo.

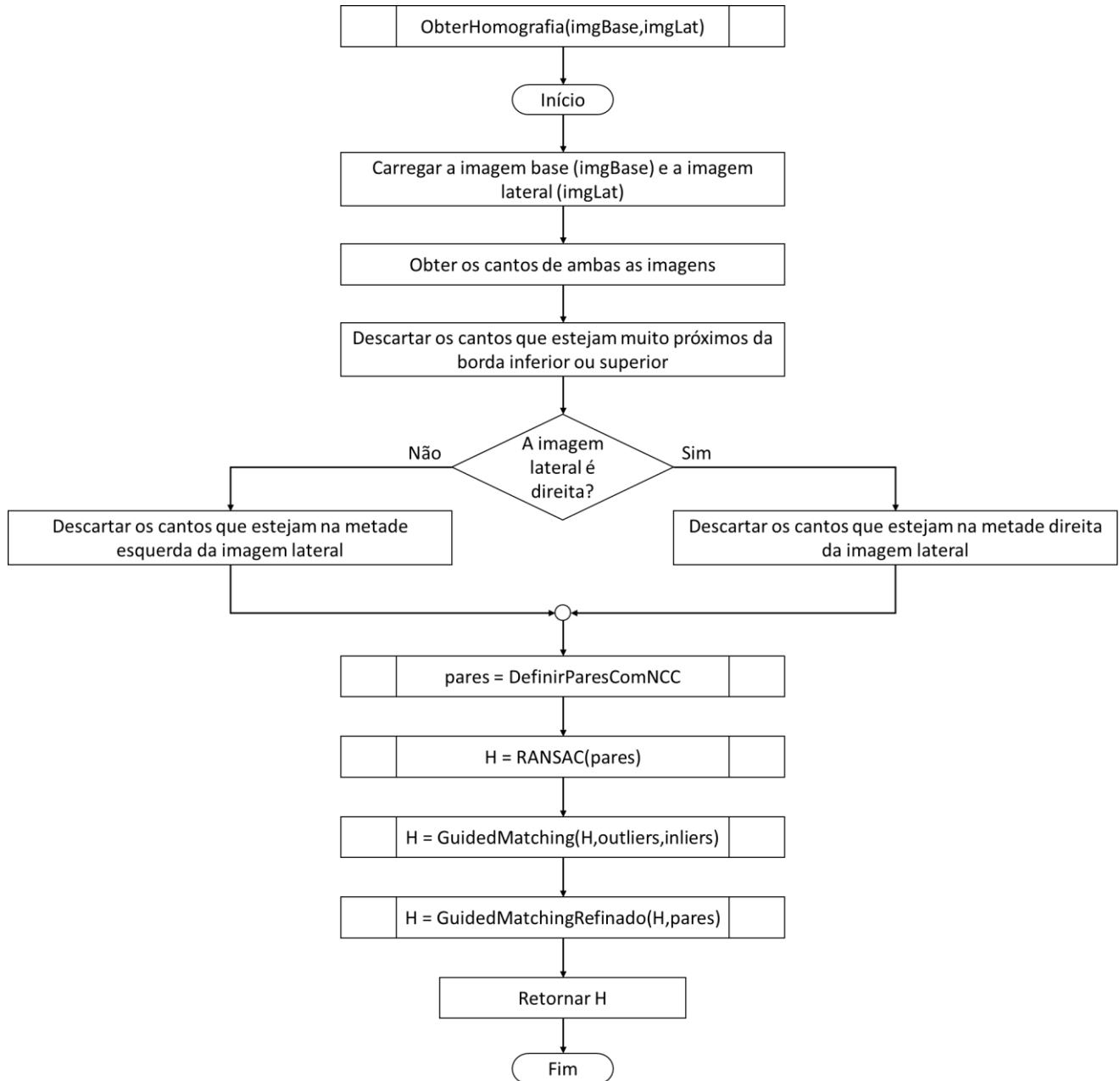


Figura 7.23 Fluxograma da sub-rotina de obtenção da homografia

O tempo médio de processamento da obtenção das quatro homografias para esse exemplo é de 47 segundos.

8. CONSTRUÇÃO DOS VÍDEOS

A construção dos vídeos é feita utilizando um programa desenvolvido em linguagem C++ com a biblioteca OpenCV, que faz a aplicação da transformação de mosaico quadro a quadro em cada imagem. Após feitos os panoramas de cada olho, os mesmos são colocados no formato *Stereo Over Under*, formando um novo quadro do vídeo final.

O programa utiliza os vídeos cortados pelo programa de aquisição e as homografias obtidas para montar o vídeo *Stereo Over Under*.

O tempo de processamento (Tp) necessário para fazer a construção de um vídeo *Stereo Over Under* com quantidade de quadros igual a Nf a partir dos vídeos cortados é apresentado na Tabela 8.1.

Tabela 8.1 Tempo de processamento da criação de vídeos

Tamanho do vídeo	Nf	Tp (s)	Tp/Nf
640 x 480	57	125	2,193
640 x 480	56	119	2,125
640 x 480	97	203	2,093
640 x 480	125	250	2,000

Na Tabela 8.1 é possível observar que, em média, são necessários 2 segundos de processamento para cada quadro, considerando quadros com formato 640 x 480.

O fluxograma que ilustra o processo é apresentado na Figura 8.1:

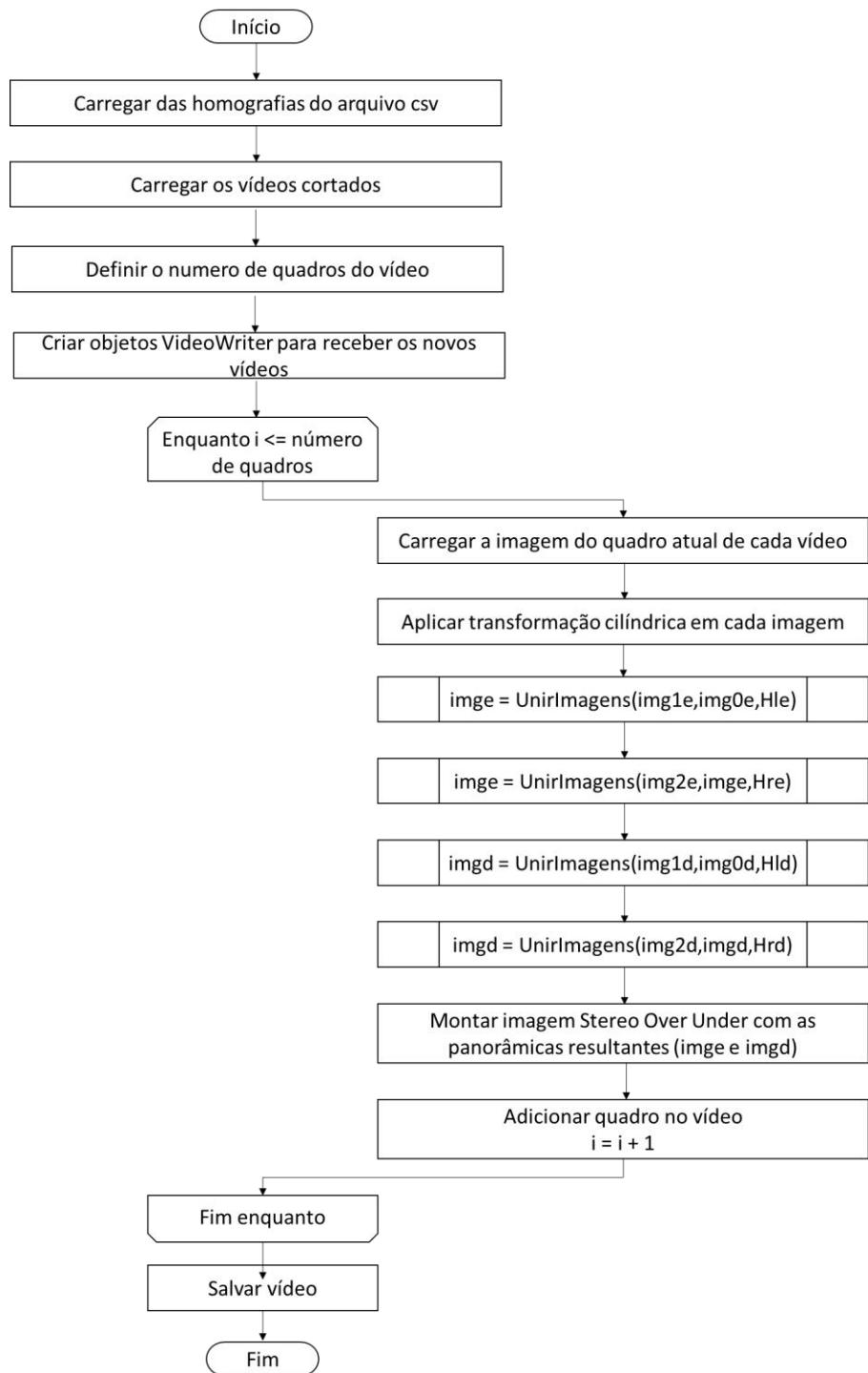


Figura 8.1 Fluxograma da criação de vídeos

9. REPRODUÇÃO DOS VÍDEOS

Após prontos, os vídeos devem ser reproduzidos de modo que um trecho da panorâmica esquerda seja projetado para o olho esquerdo e um trecho da panorâmica direita seja projetada para o olho direito, garantindo que os efeitos de 360° e 3D se mantenham e sejam percebidos.

A seleção do trecho da imagem panorâmica a ser exibido pelo aplicativo é feita por meio da movimentação da cabeça do usuário. O trecho da imagem visto pelo usuário é selecionado por meio de sensores de orientação presentes no smartphone, assim, o software é capaz de compor uma imagem nova para cada olho (Figura 9.1).

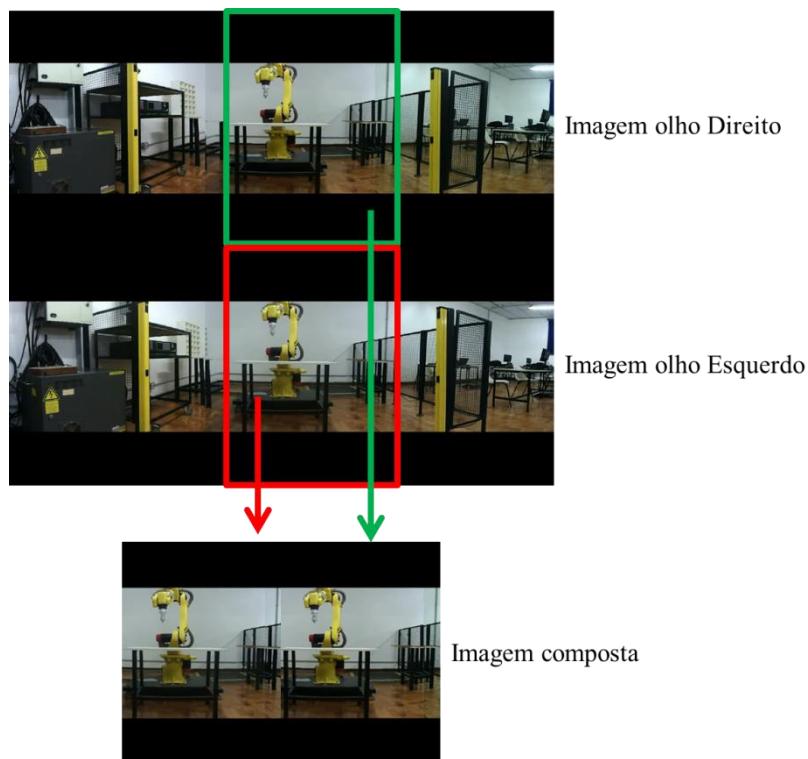


Figura 9.1 Projeção do vídeo

9.1.1. Programa de projeção

No projeto são utilizados vídeos no formato *Stereo Over Under* onde a imagem panorâmica superior é criada pela junção das imagens das câmeras de olho direito de cada módulo e a imagem panorâmica inferior é criada pela junção das imagens das câmeras de olho esquerdo de cada módulo.

O aplicativo que faz a projeção foi desenvolvido em Java a partir do Android Studio. A escolha do vídeo a ser projetado é feita pelo usuário através de uma lista com todos os programas presentes na pasta Movie do celular (Figura 9.2).

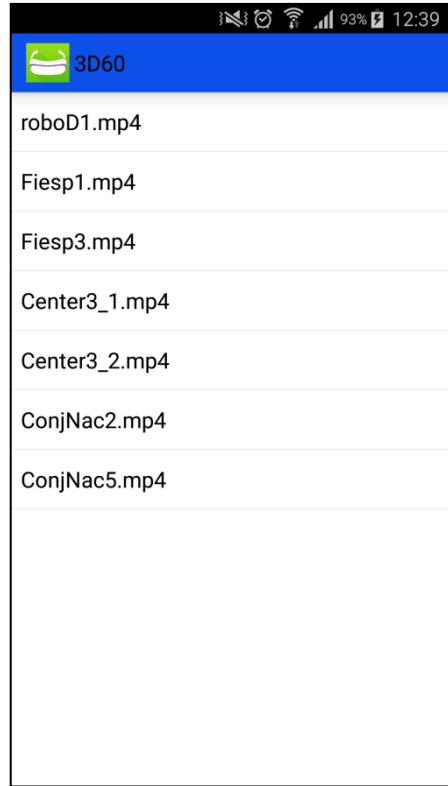


Figura 9.2 Tela do aplicativo com a lista de vídeos

O aplicativo utiliza a classe `VrVideoView`, do Android Studio, para projetar o vídeo panorâmico de modo a gerar o efeito de 3D, conforme Figura 9.1 e também faz a compensação da deformação causada pelas lentes do óculos, como visto na Figura 9.3.

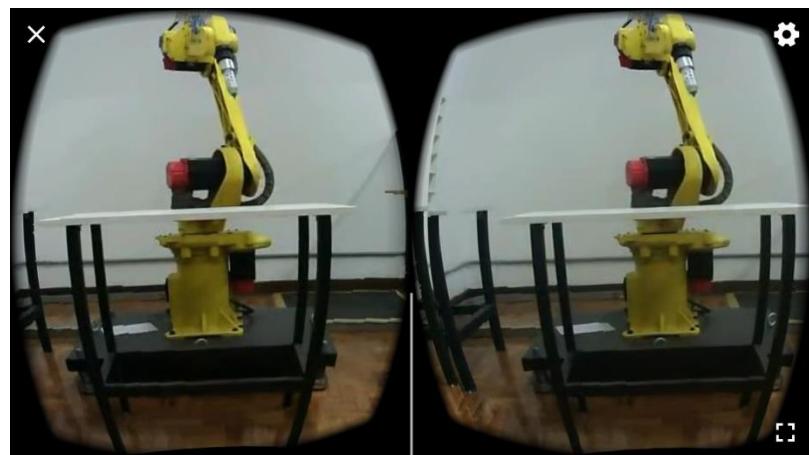


Figura 9.3 Tela do aplicativo projetando o vídeo

9.1.2. Fluxograma do programa de projeção do vídeo

A lógica utilizada no aplicativo segue o fluxograma mostrado na Figura 9.4. O aplicativo continua em funcionamento até que o usuário o feche, quando um arquivo da lista é selecionado o vídeo é aberto no VrVideoView e projetado como um vídeo em 360° para que a projeção do vídeo seja feita em 360° e 3D é necessário abrir o modo óculos VR do VrVideoView.

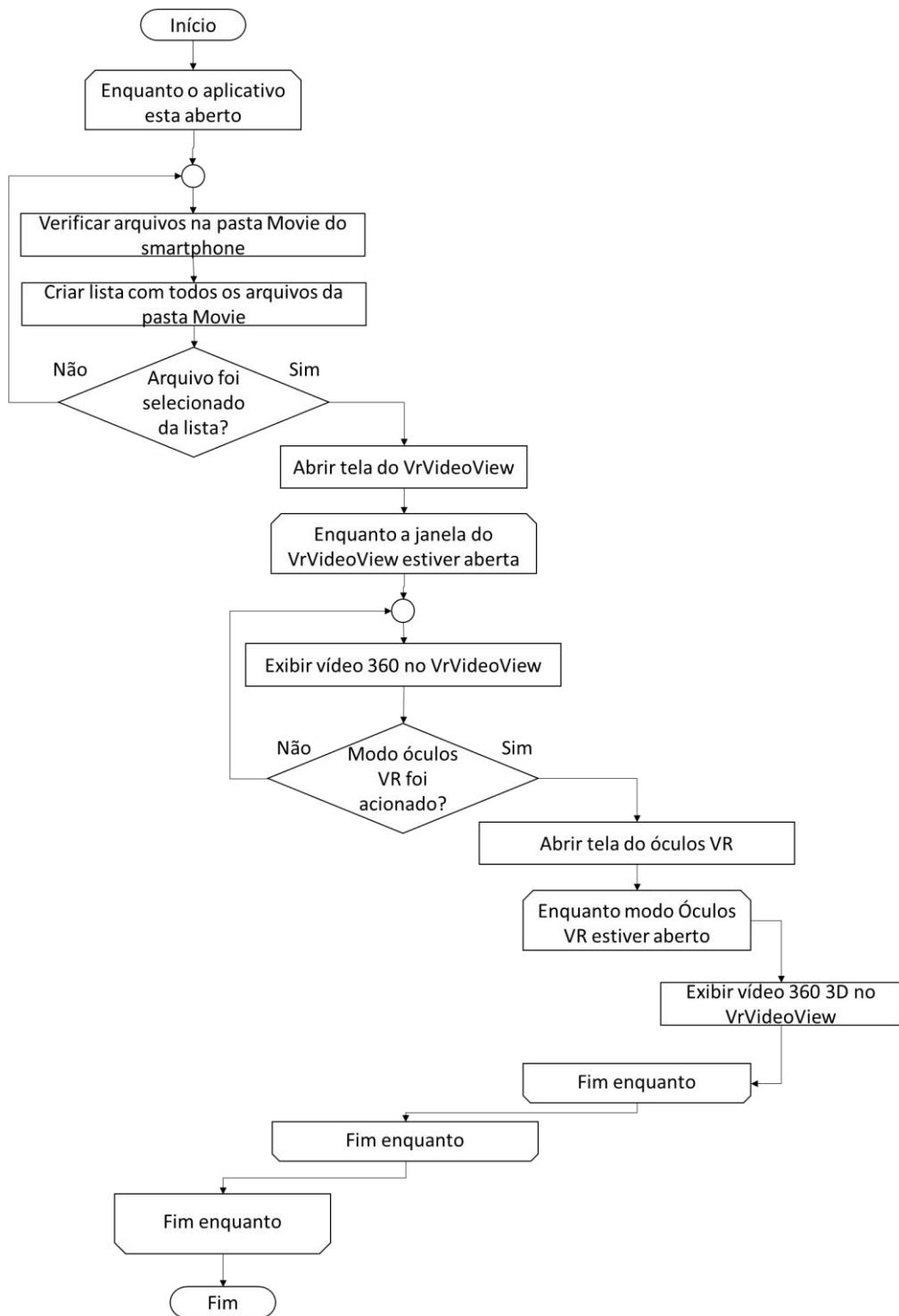


Figura 9.4 Fluxograma do programa de projeção

9.1.3. Óculos VR

A visualização dos vídeos em 3D necessita de um equipamento especial para visualização de imagens em 3D pelo método de estereoscopia.

Uma alternativa é utilizar óculos de realidade virtual, como o Google Cardboard (Figura 9.5) ou o VR Box (Figura 9.6). Ambos consistem em um suporte no qual é possível colocar um smartphone com sistema operacional Android e visualizar sua tela dentro do óculos.



*Figura 9.5 Google Cardboard
Fonte: Google (2017).*



Figura 9.6 Óculos VR Box

Esses tipos de óculos vêm equipados com uma lente biconvexa para cada olho, permitindo que o usuário tenha uma sensação de imersão ao assistir um vídeo em 3D. Em seu interior existe uma separação entre os olhos, para que a imagem projetada para cada olho não seja percebida pelo outro gerando a sensação de profundidade.

Para realizar os testes de filmagem da câmera é necessário utilizar um óculos de realidade virtual nesse projeto é utilizado o óculos VR Box.

10. RECEPÇÃO DO PÚBLICO

Durante a feira de exposição pública dos trabalhos de TCC do Instituto Mauá de Tecnologia, a Eureka 2017, foi feita uma pesquisa de mercado para saber qual seria o público alvo caso no futuro o projeto vire um produto para comercialização. Os resultados foram obtidos de uma amostra de 256 pessoas.

Verificou-se que 54% das pessoas físicas que visitaram o estande de exposição do trabalho, compraria a câmera se custasse até R\$ 1.000,00, preço abaixo dos custos necessários para sua fabricação. Apenas 4% compraria independente do preço de venda. (Figura 10.1)

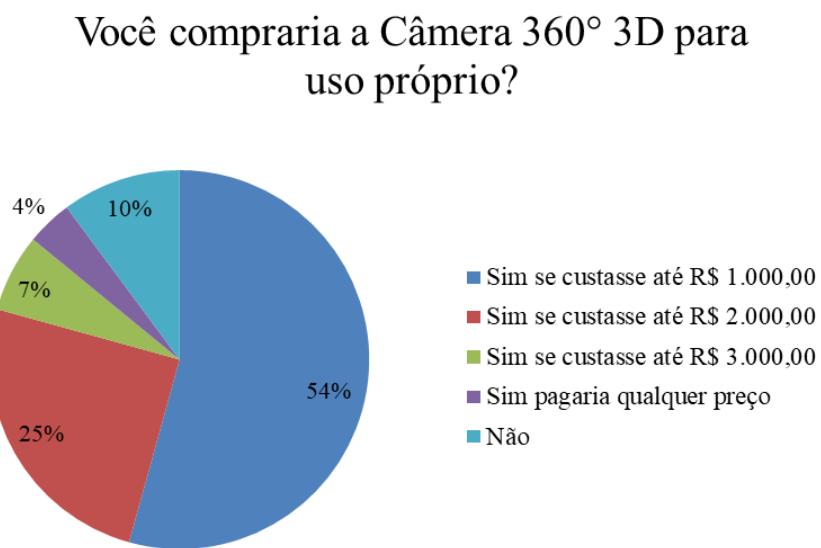


Figura 10.1 Resultado da pergunta da venda da câmera

O público alvo possível para aquisição da câmera, seria uma empresa de eventos. Tanto de eventos menores como casamentos e festas de debutantes, como um grande show de um artista conhecido. Essa conclusão foi possível a partir das perguntas as pessoas “Supondo que tivesse que contratar uma empresa de eventos para filmagem, contrataria uma com esse produto?” e “Você gostaria de ver um vídeo 360° 3D de um show?”, que mostrou que 95% das pessoas contrataria uma empresa de eventos que possua essa tecnologia em sua gama de produtos (Figura 10.2) e 95% também gostaria de assistir um show em 360° 3D (Figura 10.3).

Supondo que tivesse que contratar um
empresa de eventos para filmagem,
contrataria uma com esse produto?

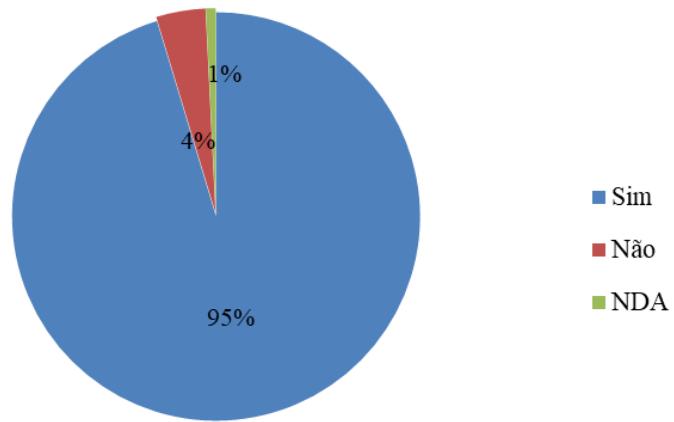


Figura 10.2 Resultado da pergunta de venda de serviços

Você gostaria de ver um vídeo 360° 3D
de um show?

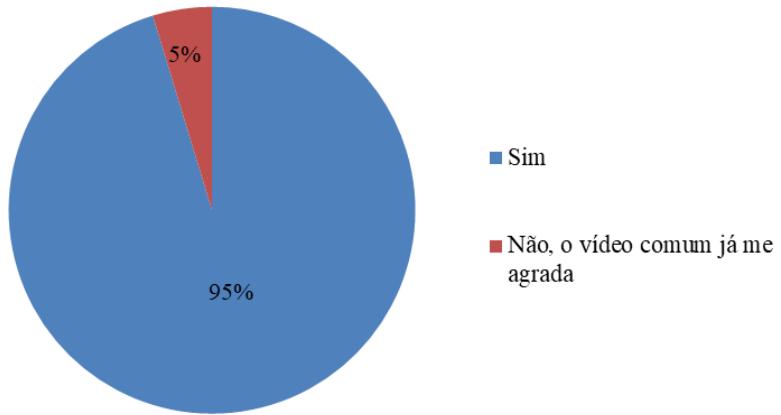


Figura 10.3 Resultado da pergunta de sublocação de serviços

Entende-se, portanto, que há maior oportunidade na utilização da câmera para venda de serviços de filmagens em relação à venda da câmera como dispositivo de uso pessoal.

11. CONCLUSÃO

A câmera 360° 3D obteve um desempenho adequado durante os testes e boa aceitação do público durante a Eureka 2017. Vídeos gravados com ela, montados e reproduzidos por meio das técnicas descritas neste trabalho, puderam ser vistos por mais de uma centena de espectadores que conseguiram perceber os dois efeitos na gravação.

Em termos de praticidade um ponto a salientar é a impossibilidade de se fazer a gravação e a reprodução em tempo real, por conta do alto volume de processamento necessário em cada etapa do processo. Esse problema poderia ser contornado com a criação de um hardware específico para a aplicação.

Em relação à presença de diferentes planos de cena, ficou evidente que as técnicas de mosaico aqui utilizadas funcionam apenas para o plano no qual a maior parte dos pontos notáveis (cantos) estava presente. Planos mais próximos ou mais afastados apresentam problemas de junção.

A concepção, projeto, criação, teste e validação da câmera 360° 3D foram atividades que trouxeram um aprendizado inestimável ao grupo e permitiu a utilização da engenharia como meio de criação e desenvolvimento de tecnologias.

11.1. Sugestões para trabalhos futuros

A melhoria da qualidade das imagens e o panorama completo auxiliam na melhoria da sensação de imersão, portanto, melhores resultados podem ser obtidos ao utilizar câmeras de maior resolução e que se faça a filmagem de todos os 360°.

A utilização de algoritmos de ajuste de cor e brilho auxilia na correção de problemas relacionados à diferença de exposição que cada câmera sofre, equilibrando as cores e melhorando a percepção do usuário.

Para garantir que os mosaicos tenham a melhor qualidade e para dar autonomia à câmera deve ser desenvolvido hardware adequado que consiga processar as imagens capturadas em tempo real sem a necessidade de utilizar um computador.

12. REFERÊNCIAS

360Rize. **360eros and Everest team are awarded Guinness Book of World Records.** Olean, 2014, Disponível em: <<http://360heros.advantinteractive.com/2014/01/360heros-and-everest-team-are-awarded-guinness-book-of-world-records-certificate-2/>> Acesso em: 14 abr. 2017.

360Rize. **360Heros.** Disponível em: <<https://www.360rize.com/>> Acesso em: 14 abr. 2017.

BLOCK, B. **3D Storytelling:** How Stereoscopic 3D Works and How to Use It. Burlington: Focal Press, 2013.

Boris Smus. **Three approaches to VR lens distortion.** Boris Smus, Canada, 20 abr. 2016. Disponível em: <<http://smus.com/vr-lens-distortion/>> Acesso em: 08 ago. 2017.

BRADLEY, A. **Curved TV screens: what's all the hype about?** Cobra simulation. United Kingdom, 13 mai. 2014. Disponível em: <<http://www.cobrasimulation.com/about-cobra/news/blog/category/may-2014>> Acesso em: 08 ago. 2017.

BRIECHLE, K.; HANEBECK, U. D. **Template Matching using Fast Normalized Cross Correlation.** München, Germany: Institute of Automatic Control Engineering, Technische Universität München, 2001. 8p.

CALEIRO, J. P. **3 sinais que a indústria do entretenimento mudou para valer.** Exame.com. São Paulo, 21 mar. 2015. Disponível em: <<http://exame.abril.com.br/economia/3-sinais-que-a-industria-do-entretenimento-mudou-pra-valer>> Acesso em: 11 abr. 2017.

CAPEL, D. P. **Image Mosaicing and Super-resolution.** Oxford: University of Oxford, 2001. 263p. Disponível em: <<https://www.robots.ox.ac.uk/~vgg/publications/2001/Capel01a/capel01a.pdf>> Acesso em 18 mar. 2017.

COSSETTI, M. C. **YouTube agora suporta vídeos em 360 graus; veja como funciona.** Techtudo, 17 ago. 2015. Disponível em: <<http://www.techtudo.com.br/noticias/noticia/2015/03/youtube-agora-suporta-videos-em-360-graus-veja-como-funciona.html>> Acesso em: 16 abr. 2017.

FISCHLER, M. A., BOLLES, R. C. **Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.** Comm Assoc. Com Mach, 1981.

Google. **com.google.vr.sdk.widgets.videos.** GoogleVR. 1 fev. 2017. Disponível em: <<https://developers.google.com/vr/android/reference/com/google/vr/sdk/widgets/video/package-summary>> Acesso em: 07 ago. 2017.

Google. **Google Cardboard.** Palo Alto, 2017. Disponível em: <<https://vr.google.com/cardboard>> Acesso em: 15 abr. 2017.

Google. **Google VR Java API Overview.** GoogleVR. 7 dez. 2016. Disponível em: <https://developers.google.com/vr/android/reference_overview> Acesso em: 07 ago. 2017.

HARRIS, C.; STEPHENS, M. **A combined corner and edge detector.** United Kingdom: Plessey Research Roke Manor, 1988. 5p.

HumanEyes Technologies. **Vuze Camera: About us.** Disponível em: <<http://vuze.camera/about-us/>> Acesso em: 14 abr. 2017.

LANDIM, W. **Como funcionam os diferentes tipos de 3D?** Tecmundo, 01 fev. 2011. Disponível em: <<https://www.tecmundo.com.br/3d/8154-como-funcionam-os-diferentes-tipos-de-3d-.htm>> Acesso em: 17 abr. 2017

MARCHESAN, R. **Investimento em publicidade online no país cresce 26% em 2016, diz pesquisa.** UOL, São Paulo, 29 mar. 2017. Disponível em:

<<https://economia.uol.com.br/noticias/redacao/2017/03/29/investimento-em-publicidade-online-no-pais-cresce-26-em-2016-diz-pesquisa.htm>> Acesso em: 15 abr. 2017.

OLIVEIRA, F. Entretenimento e mídia vão crescer mais no Brasil que no mundo, diz PwC.
Folha de São Paulo, São Paulo, 14 jun. 2016. Disponível em:
<<http://www1.folha.uol.com.br/mercado/2016/06/1781343-entretenimento-e-midia-vao-crescer-mais-no-brasil-que-no-mundo-diz-pwc.shtml>>. Acesso em 11 abr. 2017.

OPENCV. OpenCV 2.4.13.2 documentation. OpenCV, 2016. Disponível em:
<<http://docs.opencv.org/2.4.13.2/>>. Acesso em: 14 ago. 2017.

Panocam 3D. **Panocam 3D.** Disponível em: <<http://www.panocam3d.com/>> Acesso em: 14 abr. 2017.

Rede Paranaense de Comunicação. **O show não pode parar: mercado de entretenimento está otimista.** In: De olho no mercado. Curitiba, 26 out. 2015. Disponível em:
<<http://www.negociosrpc.com.br/deolhonomercado/economia/o-show-nao-pode-parar-mercado-de-entretenimento-esta-otimista/>> Acesso em: 15 abr. 2017.

SOARES, R. O ano do vídeo online. Meio & Mensagem, 29 mar. 2017. Disponível em:
<<http://www.meioemensagem.com.br/home/opiniao/2017/03/29/o-ano-do-video-online.html>>
Acesso em: 15 abr. 2017.

SZELISKY, R. Computer Vision: Algorithms and Applications. Springer, 2010. 972p.
Disponível em: <<http://szeliski.org/Book/>>. Acesso em: 18 mar. 2017.

Time For Fun. **Setor de entretenimento no Brasil.** In: Visão Geral do Setor na América do Sul e Oportunidades de Crescimento. São Paulo, 01 fev. 2017. Disponível em:
<http://ri.t4f.com.br/timeforfun/web/conteudo_pt.asp?idioma=0&conta=28&tipo=34927>.
Acesso em: 11 abr. 2017.

ANEXOS

A. FLUXOGRAMA DAS SUBROTINAS DE OBTEÇÃO DA HOMOGRAFIA

Fluxograma da obtenção de pares utilizando o método NCC

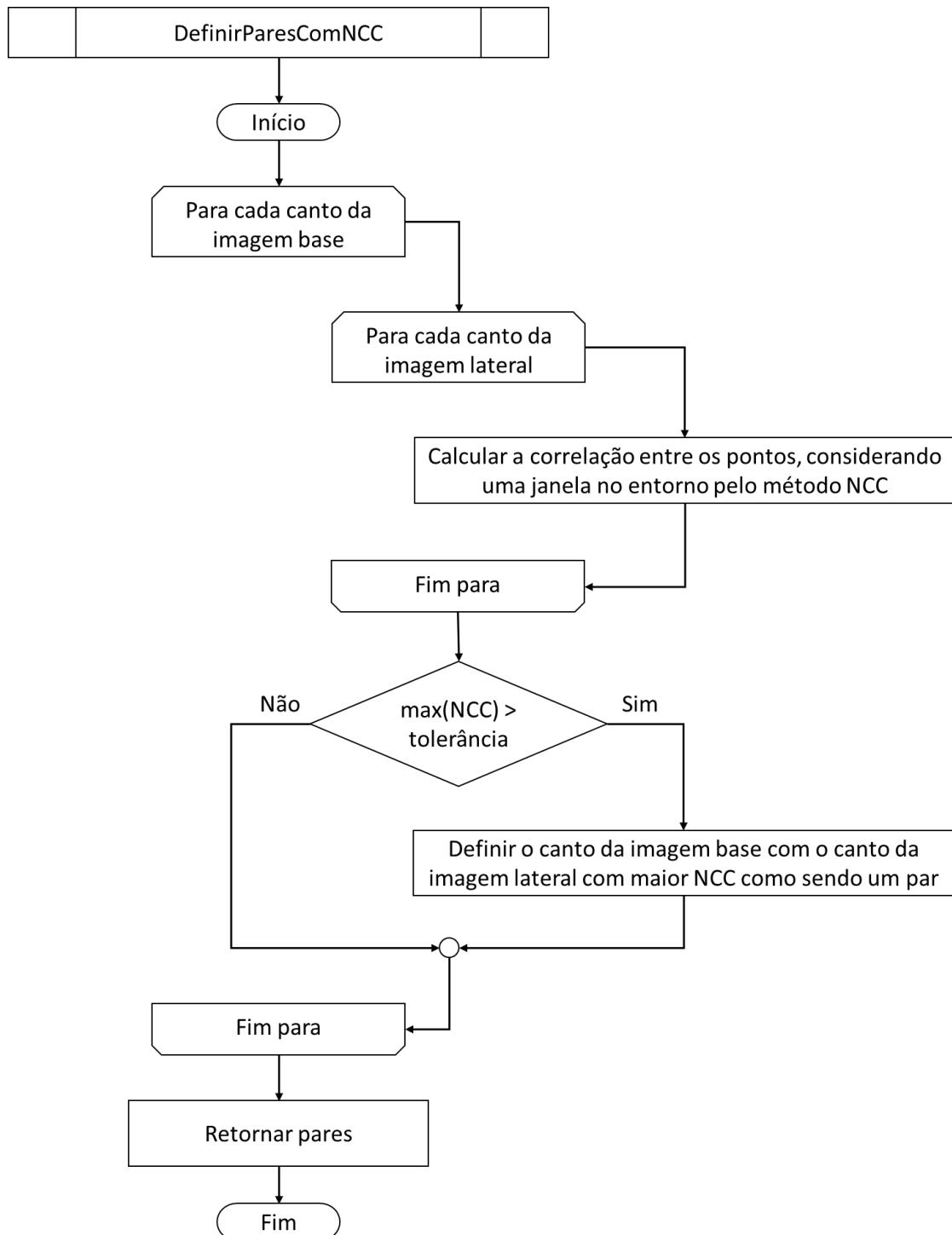


Figura A.1 Obtenção de pares por NCC

Fluxograma da obtenção de homografia utilizando o método RANSAC

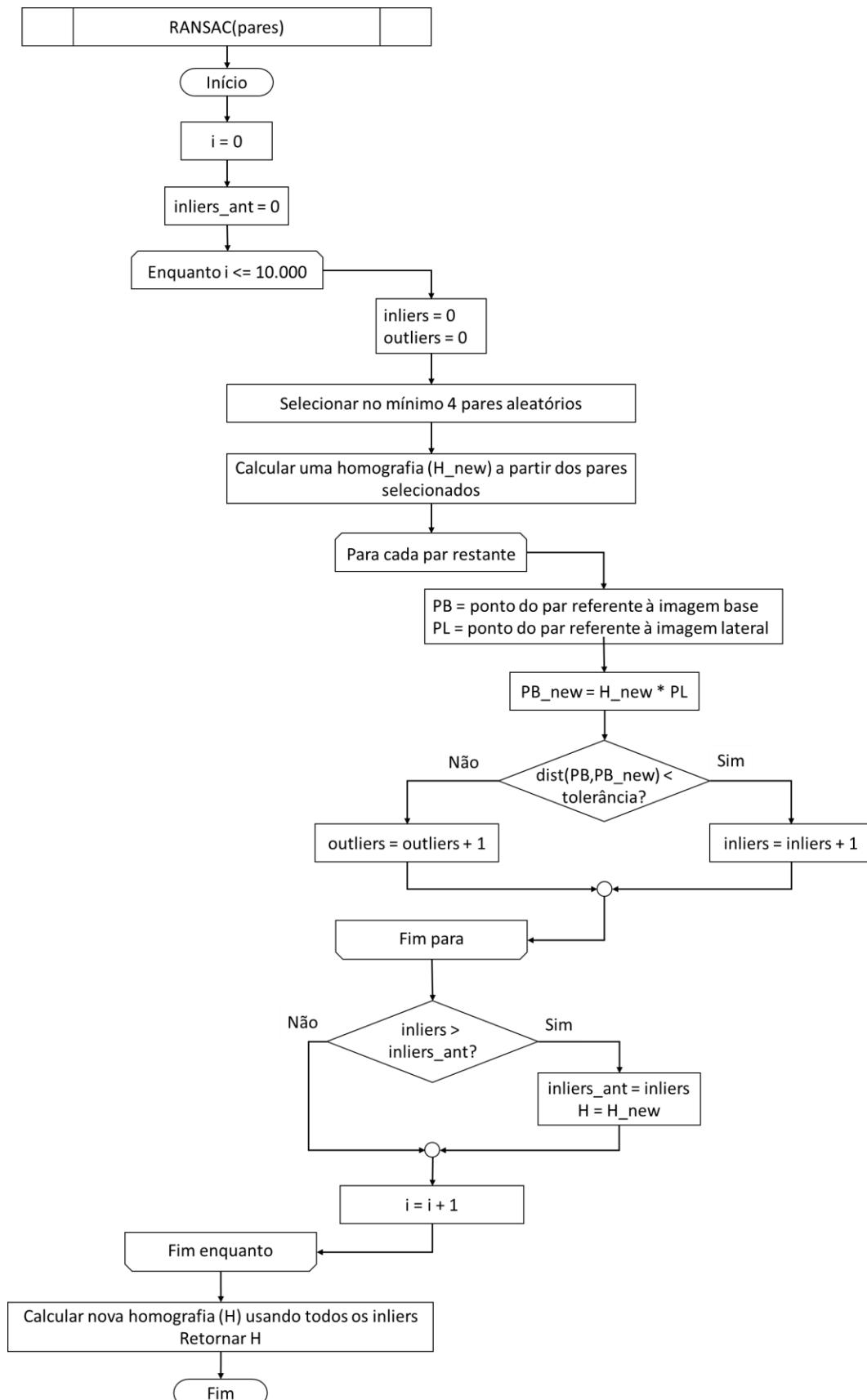


Figura A.2 Obtenção de homografia pelo método RANSAC

Fluxograma da obtenção de homografia utilizando o método Guided Matching

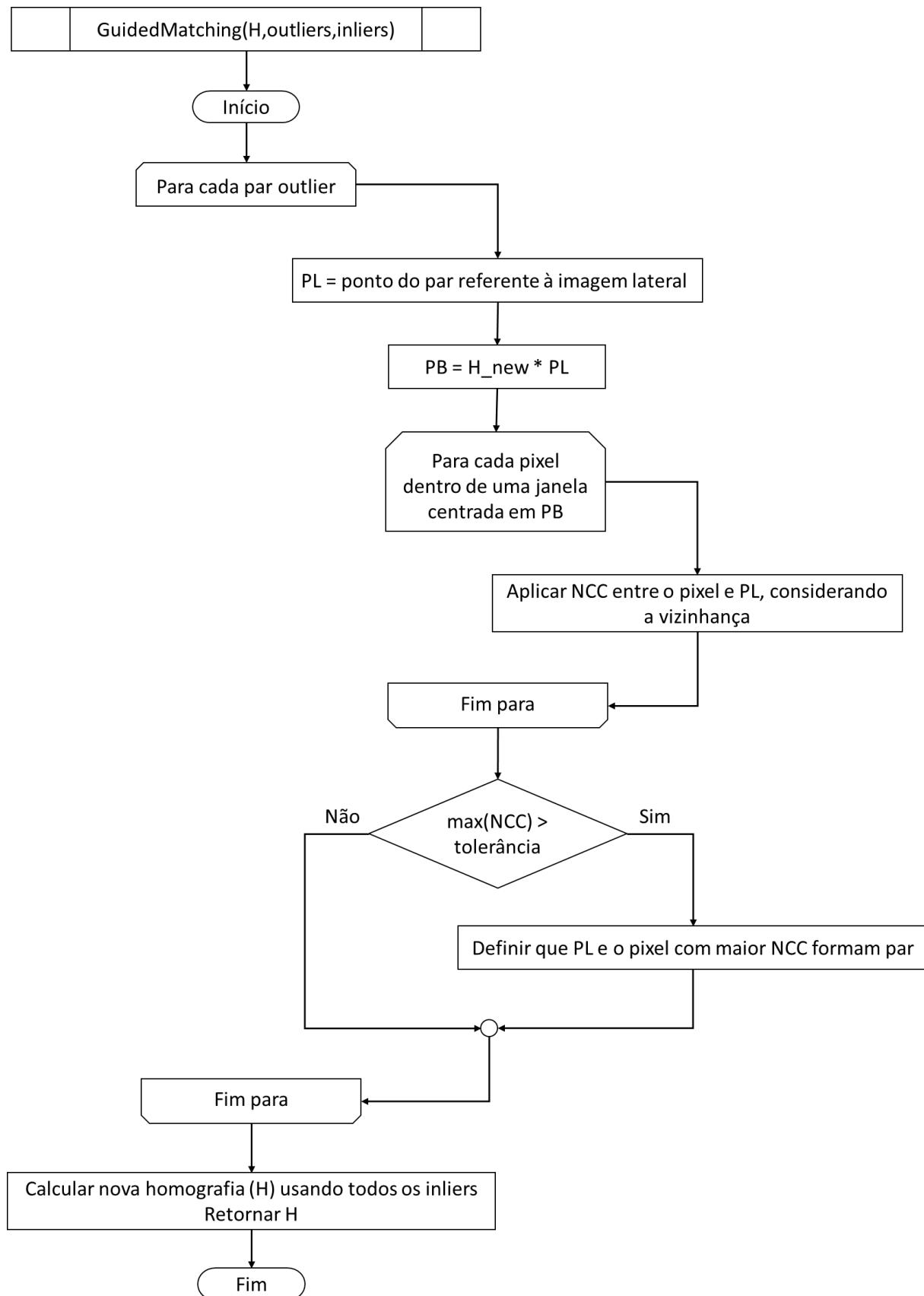


Figura A.3 Obtenção de homografia pelo método Guided Matching

Fluxograma da obtenção de homografia utilizando o método Guided Matching Refinado

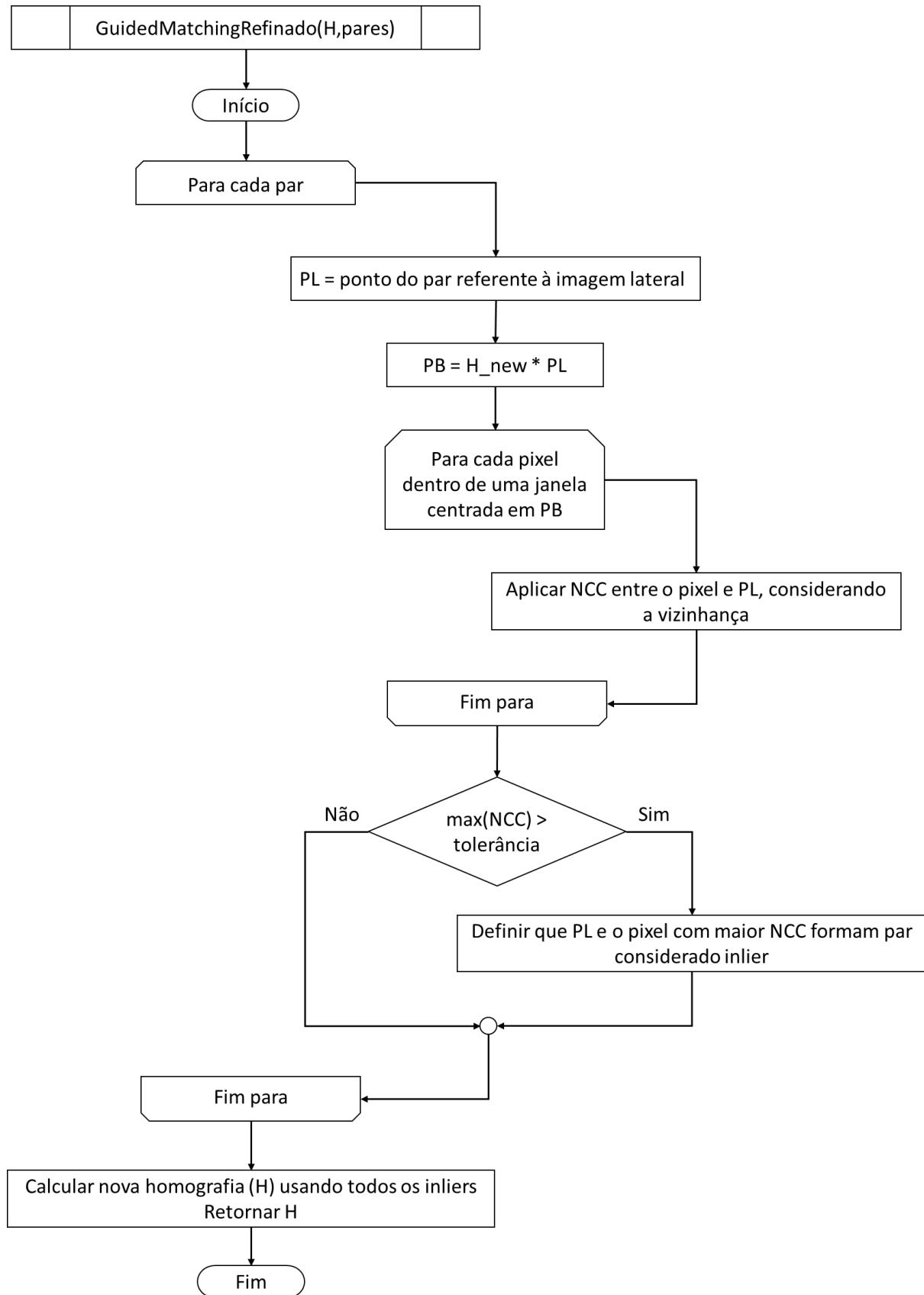


Figura A.4 Obtenção de homografia pelo método Guided Matching Refinado